

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2004, 11:06:56 ; Search time 21.0943 Seconds
(without alignments)
3126.121 Million cell updates/sec

Title: US-10-658-989a-1

Perfect score: 1184
Sequence: 1 GPGPEPTGLPGPGERGG.....GEQGVCDLGAPGSPAGG 209

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mmc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1149	97.0	1461	4	O76045 homo sapien
2	1149	97.0	1464	4	Q8N473
3	1122	94.8	1453	11	Q63079 rattus norv
4	1116	94.3	1453	11	Q81039 mus musculu
5	989	83.5	1450	13	Q9Y1B4 cynops pyrr
6	981	82.9	1445	13	Q3251 rana catesb
7	962.5	81.3	1449	13	Q802B5 xenopus lae
8	941	79.5	1447	13	Q91B91 xenopus lae
9	860	72.6	1418	6	Q28396 equus cabal
10	860	72.6	1487	6	Q77753 canis famil
11	859	72.6	1160	4	Q14046 homo sapien
12	859	72.6	1269	13	Q7T2Z7 gallus gall
13	859	72.6	1420	13	Q90W37 gallus gall
14	859	72.6	1487	4	Q14047 homo sapien
15	850	71.8	1486	13	Q91717 xenopus lae
16	850	71.8	1486	13	Q7ZT16 xenopus lae

17	844	71.3	1491	13	Q91718
18	844	71.3	1491	13	Q7ZTM4
19	839	70.9	1419	11	Q63123
20	839	70.9	1442	11	Q82031
21	839	70.9	1442	11	Q62033
22	839	70.9	1459	11	Q62032
23	835	70.5	1419	11	Q80X38
24	835	70.5	1419	11	Q80VY3
25	833	70.4	1418	13	Q9W7B9
26	833	70.4	1449	13	Q910C0
27	798	67.4	1347	4	Q86QB3
28	776	65.5	138	4	Q9UML6
29	776	65.5	1163	4	Q8N6U4
30	771	65.1	1497	11	Q7TMS0
31	768	64.9	1458	13	Q910B9
32	765	64.6	1497	11	Q61431
33	763	64.4	998	11	Q8CFM4
34	763	64.4	1222	11	Q8K173
35	763	64.4	1464	11	Q8BK12
36	763	64.4	1464	11	Q8BLW4
37	763	64.4	1464	11	Q7T732
38	760	64.2	1314	11	Q8CGA5
39	745	62.9	1366	4	Q15177
40	745	62.9	1366	4	Q725S6
41	717.5	60.6	1258	13	Q8AW11
42	705.5	59.6	1352	13	Q90Y00
43	691.5	58.4	1352	13	Q8U0J4
44	689.5	58.2	1346	13	Q8U0J3
45	678.5	57.3	1346	13	Q801M5

ALIGNMENTS

RESULT 1

O76045 PRELIMINARY; PRT; 1461 AA.

AC O76045; 1998 (TREMREL. 08, Created)
 DT 01-NOV-1998 (TREMREL. 12, Last sequence update)
 DT 01-OCT-2003 (TREMREL. 25, Last annotation update)
 DE Pro alpha 1(I) collagen.
 GN COL1A1
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 CX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85130970; PubMed=2857713;
 RA Chu M.L., de Wet W., Bernard M., Ramirez F.;
 RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
 Promoter structure, AluI repeats, and polymorphic transcripts.";
 J. Biol. Chem. 260:2315-2320(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88329734; PubMed=2843432;
 RA D'Alessio M., Bernard M., Pretorius P.J., de Wet W., Ramirez F.;
 RT "Complete nucleotide sequence of the region encompassing the first
 twenty-five exons of the human pro alpha 1(I) collagen gene.";
 Gene 67:105-115(1988).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89025644; PubMed=3178743;
 RA Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
 Jaenisch R., Prockop D.J.;
 RT "Structure of a full-length cDNA clone for the prepro alpha 1(I) chain
 of human type I procollagen.";
 Biochem. J. 253:919-922(1988).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91138770; PubMed=1995349;
 RA Maatta A., Bornstein P., Penttinen R.P.;

RT "Highly conserved sequences in the 3'-untranslated region of the
 RT COL1A1 gene bind cell-specific nuclear proteins.";
 RL FEBS Lett. 279:9-13(1991).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92157916; PubMed=1787829;
 RA Westerhausen A., Constantinou C.D., Pack M., Peng M.Z., Hanning C.,
 RA Olsen A.S., Prockop D.J.;
 RT "Completion of the last half of the structure of the human gene for
 RT the Pro alpha 1 (I) chain of type I procollagen (COL1A1).";
 RL Matrix 11:375-379(1991).
 RN [6]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98107942; PubMed=9443882;
 RA Korhko J., Ala-Korhko L., De Paape A., Nuytinck L., Earley J.,
 RA Prockop D.J.;
 RT "Analysis of the COL1A1 and COL1A2 genes by PCR amplification and
 RT scanning by conformation-sensitive gel electrophoresis identifies only
 RT COL1A1 mutations in 15 patients with osteogenesis imperfecta type I;
 RT identification of common sequences of null-allele mutations.";
 RL Am. J. Hum. Genet. 62:98-110(1998).
 RN [7]
 RP SEQUENCE FROM N.A.
 RA Korhko J.M., Earley J.J., Nuytinck L., DePaape A., Prockop D.J.,
 RA Ala-Korhko L.;
 RL Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF017178; AAB94054.2; -
 DR GO; GO:0005581; C:collagen; IEA.
 DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
 DR InterPro; IPR008161; Clg helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR002181; Fibrinogen_C.
 DR InterPro; IPR000885; Fib_collagen_C.
 DR InterPro; IPR001007; VWF_C.
 DR Pfam; PF01410; COLFI; 1.
 DR ProDom; PD000007; Clg helix; 2.
 DR ProDom; PD002078; Fib_collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 DR SMART; SM00214; VMC; 1.
 DR PROSITE; PS01208; VWFC_1; 1.
 DR PROSITE; PS0184; VWFC_2; 1.
 KW Hypothetical protein; IEA.
 SQ SEQUENCE 1461 AA; 138630 MW; 9ACF6DE30EA78E21 CRC64;

Query Match 97.0%; Score 1149; DB 4; Length 1461;
 Best Local Similarity 93.6%; Pred. No. 8.1e-72;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GPPGPTGLPGRGPGSRGPGADGVAGPKGPAGERSGPGA 48
 DB 455 GPPGAGEGKRGARGFPTGLPGRGPGSRGPGADGVAGPKGPAGERSGPGA 514
 QY 49 GPKSGPEAGRPGAGLPGAKGLTSGSPSPDGKTPGPPAGQDGRPGPPPGARGQA 108
 DB 515 GPKSGPEAGRPGAGLPGAKGLTSGSPSPDGKTPGPPAGQDGRPGPPPGARGQA 574
 QY 109 GVMGFPKPKGAAGRPGKAGRGVPGPCAVGPKGCEAGACQPPGPPGAGRGEGGPA 168
 DB 575 GVMGFPKPKGAAGRPGKAGRGVPGPCAVGPKGCEAGACQPPGPPGAGRGEGGPA 634
 QY 169 GSPGFQGLPGPAGPPGAGKPGEGQVPGDLGAPGSPGAG 208
 DB 635 GSPGFQGLPGPAGPPGAGKPGEGQVPGDLGAPGSPGARG 674

RESULT 2
 Q8N473
 ID Q8N473 PRELIMINARY; PRT; 1464 AA.
 AC Q8N473
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Hypothetical protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RL Strausberg R.;
 RL Submitted (Aug-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC036531; AAB36531.1; -
 DR GO; GO:0005581; C:collagen; IEA.
 DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
 DR InterPro; IPR008161; Clg helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR002181; Fibrinogen_C.
 DR InterPro; IPR000885; Fib_collagen_C.
 DR InterPro; IPR001007; VWF_C.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 18.
 DR ProDom; PD000007; Clg helix; 2.
 DR ProDom; PD002078; Fib_collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 DR SMART; SM00214; VMC; 1.
 DR PROSITE; PS01208; VWFC_1; 1.
 DR PROSITE; PS0184; VWFC_2; 1.
 KW Hypothetical protein; IEA.
 SQ SEQUENCE 1464 AA; 139011 MW; B0581F8D1C89DDE8 CRC64;
 Query Match 97.0%; Score 1149; DB 4; Length 1464;
 Best Local Similarity 93.6%; Pred. No. 8.1e-72;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GPPGPTGLPGRGPGSRGPGADGVAGPKGPAGERSGPGA 48
 DB 458 GPPGAGEGKRGARGFPTGLPGRGPGSRGPGADGVAGPKGPAGERSGPGA 517
 QY 49 GPKSGPEAGRPGAGLPGAKGLTSGSPSPDGKTPGPPAGQDGRPGPPPGARGQA 108
 DB 518 GPKSGPEAGRPGAGLPGAKGLTSGSPSPDGKTPGPPAGQDGRPGPPPGARGQA 577
 QY 109 GVMGFPKPKGAAGRPGKAGRGVPGPCAVGPKGCEAGACQPPGPPGAGRGEGGPA 168
 DB 578 GVMGFPKPKGAAGRPGKAGRGVPGPCAVGPKGCEAGACQPPGPPGAGRGEGGPA 637
 QY 169 GSPGFQGLPGPAGPPGAGKPGEGQVPGDLGAPGSPGAG 208
 DB 638 GSPGFQGLPGPAGPPGAGKPGEGQVPGDLGAPGSPGARG 677

RESULT 3
 Q63079
 ID Q63079 PRELIMINARY; PRT; 1453 AA.
 AC Q63079
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Collagen alpha1 (fragment).
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE OF 1-1092 FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Tooth;
 RA Brandsten C., Lundmark C., Christerson C., Hammarstrom L., Wurtz T.;
 RT "Expression of Collagen alpha1(I) mRNA variants during Tooth and Bone
 RT Formation in the Rat.";
 RL J. Dent. Res. 0:0-0(0).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Tooth;
 Wurtz T.;

```
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; 278279; CAB01633.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002181; Fibrinogen_C.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001007; VWFC.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Clg_helix; 3.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWFC_1; 1.
DR PROSITE; PS01084; VWFC_2; 1.
KW Collagen.
FT NON TER
SQ SEQUENCE 1453 AA; 137887 MW; 56896BDC19A41D8 CRC64;

Query Match 94.8%; Score 1122; DB 11; Length 1453;
Best Local Similarity 91.4%; Pred. No. 5.9e-70;
Matches 201; Conservative 2; Mismatches 5; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPGPPGCGSGRGGPGADGVAGKPGAGERSGPGA 48
DB 447 GPPGAGEGKRGARGEPGSLPGPPGGERGGSGRGGPGADGVAGKPGSGERSGPGA 506

QY 49 GPKSGPAGRPGEAGLPGAKGLTSGPSGPDGKTGPPGAGQDGRPPGPPGARGQA 108
DB 507 GPKSGPAGRPGEAGLPGAKGLTSGPSGPDGKTGPPGAGQDGRPPGPPGARGQA 566

QY 109 GVMGPPGKGAAGEPKAGRGVPPGAVGPAKDGAGAGQGPAGPAGERGEGQA 168
DB 567 GVMGPPGKGTAGEPKAGRGVPPGAVGPAKDGAGAGQGPAGPAGERGEGQA 626

QY 169 GSPGFGQLPGPAGPGEAGKPGEGQVPGDLGAPGSPGAG 208
DB 627 GSPGFGQLPGPAGPGEAGKPGEGQVPGDLGAPGSPGARG 666

RESULT 4
Q810J9 PRELIMINARY; PRT; 1453 AA.
AC Q810J9
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Colon;
RA Strausberg R.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; SC050014; RAH50014.1; -.
DR GO; GO:0005581; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002181; Fibrinogen_C.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001007; VWFC.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWFC_1; 1.
DR PROSITE; PS01084; VWFC_2; 1.

QY 169 GSPGFGQLPGPAGPGEAGKPGEGQVPGDLGAPGSPGAG 208
DB 627 GSPGFGQLPGPAGPGEAGKPGEGQVPGDLGAPGSPGARG 666

RESULT 5
Q9YIB4 PRELIMINARY; PRT; 1450 AA.
AC Q9YIB4
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Alpha 1 type I collagen.
OS Cynops pyrrhogaster (Japanese common newt).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandroidae; Salamandridae; Cynops.
OX NCBI_TaxID=8330;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Regenerate forelimbs;
RX MEDLINE=99407244; PubMed=10474166;
RA Asahina K., Ohara M., Yoshizato K.;
RT "Expression of genes of type I and type II collagen in the formation
RT and development of the blastema of regenerating newt limb.";
RL Dev. Dyn. 216:59-71(1999).
DR EMBL; AB015438; BAA36973.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001007; VWFC.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Clg_helix; 1.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWFC_1; 1.
DR PROSITE; PS01084; VWFC_2; 1.
KW Collagen.
SQ SEQUENCE 1450 AA; 137564 MW; ABF8A74841B97B7C CRC64;

Query Match 83.5%; Score 989; DB 13; Length 1450;
Best Local Similarity 80.5%; Pred. No. 9.5e-61;
Matches 177; Conservative 7; Mismatches 24; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPGPPGCGSGRGGPGADGVAGKPGAGERSGPGA 48
DB 444 GPPGAGEGKRGARGEPGSLPGPPGGERGGSGRGGPGADGVAGKPGAGERSGPGA 503

QY 49 GPKSGPAGRPGEAGLPGAKGLTSGPSGPDGKTGPPGAGQDGRPPGPPGARGQA 108
DB 504 GPKSGTGESRGPGEPLPGAKGLTSGPSGPDGKTGPPGAGQDGRPPGPPGARGQA 563
```


DR GO; GO:0005581; C:collagen; IEA.
 DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
 DR InterPro; IPR008161; Clg helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR000885; Fib_collagen_C.
 DR InterPro; IPR001007; VWF_C.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF00093; vwc; 1.
 DR Pfam; PF00007; Clg helix; 2.
 DR ProDom; PD002078; Fib_collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS01208; VWF_C_1; 1.
 DR PROSITE; PS01084; VWF_C_2; 1.
 KW Collagen.
 SQ SEQUENCE 1447 AA; 137446 MW; A46DD2B4158E38B CRC64;
 Query Match 79.5%; Score 941; DB 13; Length 1447;
 Best Local Similarity 78.5%; Pred. No. 2e-57;
 Matches 172; Conservative 8; Mismatches 25; Indels 14; Gaps 2;
 QY 1 GPPGSPG-----PTGLPGPPGRRGGSGRFGADGVAGPKGAPGERGSGPGAG 49
 DB 444 GPPGSGEGKSGRGEPPAGPPGAPGERGGSGRFGSDGASGPKGAPGERGVPVGPAG 503
 QY 50 PKGSGEAGRGEAGLPKAGLGTGSPGSPGDPGKTGPPGAGQDGRPPGPPGARGQAG 109
 DB 504 PKGSGESGRGPEPLPKAGLGTGSPGSGDGKTGAPAGQDGRAGPPGPPGARGQSG 563
 QY 110 VMGPPGPKGAGEPKGAGRGVPGGAVGAGPKDGEAGAGPPGAPGAPGERGEGQGPAG 169
 DB 564 VMGFFCPKAGCEPKGKEKGVAGPPGAVGLPGKDGAGAGPPGAPGAPGERGEGQGPAG 623
 QY 170 SFGQGLPAPGPPGAPGAPGEGQGVPGDLGAPGSGPGAG 208
 DB 624 GPGFQGLPGSPGAPGSGKPGEGQGVPGDV---GPGSGPAG 659
 RESULT 9
 Q28396
 ID Q28396 PRELIMINARY; PRT; 1418 AA.
 AC Q28396;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Type II collagen.
 OS Equus caballus (Horse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
 OX NCBI_TaxID=9796;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Richardson D.W.; Dodge G.R.;
 RA Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RP SEQUENCE OF 18-68 FROM N.A.
 RA MacLeod J.N.; Fubini S.L.; Gu D.N.; Tetraault J.W.; Todhunter R.J.;
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
 RL EMBL; U62528; AAB05773.1; -.
 DR EMBL; AF040638; AAB96768.1; -.
 DR PIR; T45467; T45467.
 DR GO; GO:0005581; C:collagen; IEA.
 DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
 DR InterPro; IPR008161; Clg helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR000885; Fib_collagen_C.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 18.
 DR ProDom; PD000007; Clg helix; 4.
 DR ProDom; PD002078; Fib_collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 KW Collagen.

SQ SEQUENCE 1418 AA; 134343 MW; 115FCD19EB8696A3 CRC64;
 Query Match 72.6%; Score 860; DB 6; Length 1418;
 Best Local Similarity 74.0%; Pred. No. 7.9e-52;
 Matches 154; Conservative 10; Mismatches 44; Indels 0; Gaps 0;
 QY 1 GPFGEFGPTGLPGPPGRRGGSGRFGADGVAGPKGAPGERGSGPGAGRP 60
 DB 423 GARGEPGAGPVGPPGERRGAPGNRGFPQDGLAGPKGAPGERGSGLAGPKGANGDPGRP 482
 QY 61 GEAGLPKAGLGTGSPGSPGDPGKTGPPGAGQDGRGPPGPPGARGQAGVMGFFGPKGAA 120
 DB 483 GEPGLFOALGTGTPDAGPQGVGSGAPGEDGRRFPPOGARGQCPGVMGFFGPKGAN 542
 QY 121 GEPGKAGERGVPPGAVGAPGAGKDGSEAGAGQPPGAPGAPGERGEGQGPAGSPQFGLPGPA 180
 DB 543 GEPKAGEXGLPGAPGLRGLPGKDGTTGAAGPPGAPGAPGERGEGQGPAGSPQFGLPGPP 602
 QY 181 GPGSEAGKSGEGVPGDLGAPGSPGAG 208
 DB 603 GPPGEGKPGDQGVPGGAGAPGLVGRG 630
 RESULT 10
 O77753
 ID O77753 PRELIMINARY; PRT; 1487 AA.
 AC O77753;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Type IIA procollagen.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98340920; PubMed=9676231;
 RA Du F.; Acland G.M.; Ray J.;
 RT "Differential splicing of type II procollagen mRNA in canine retina."
 RL Anim. Biotechnol. 9:15-20 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20480698; PubMed=11024291;
 RA Du F.; Acland G.M.; Ray J.;
 RT "Cloning and expression of type II collagen mRNA: evaluation as a
 RT candidate for canine oculo-skeletal dysplasia."
 RL Gene 255:307-316 (2000).
 DR EMBL; AF023169; AAC62178.2; -.
 DR GO; GO:0005581; C:collagen; IEA.
 DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
 DR InterPro; IPR008161; Clg helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR000885; Fib_collagen_C.
 DR InterPro; IPR001007; VWF_C.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 18.
 DR Pfam; PF00093; vwc; 1.
 DR ProDom; PD000007; Clg helix; 4.
 DR ProDom; PD002078; Fib_collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS01208; VWF_C_1; 1.
 DR PROSITE; PS01084; VWF_C_2; 1.
 KW Collagen.
 SQ SEQUENCE 1487 AA; 141875 MW; 25873EAC1E311DB8 CRC64;
 Query Match 72.6%; Score 860; DB 6; Length 1487;
 Best Local Similarity 74.0%; Pred. No. 8.2e-52;
 Matches 154; Conservative 10; Mismatches 44; Indels 0; Gaps 0;
 QY 1 GPFGEFGPTGLPGPPGRRGGSGRFGADGVAGPKGAPGERGSGPGAGRP 60

```

Db 432 GARGEPGAGVGPVGPGRGAGFNGRFGQDGLAGPKGAPGERGSGLAGPKGANGDPRP 551
QY 61 GEAGLPGAGKLTGSPGSPGDKTGPAGQDGRPPGPPGARGQAGVGMFPKGA 120
Db 552 GEPGLPGAGLTGTPDAGPGKGVGSPGAPCEDGRPPGPGQARGQGVGMFPKGAN 611
QY 121 GEPKAGRGVGPVGPVGPAGVCPAGKAGGAGAGQAPGPPGAPGAGRGGQAGSPGQGLPGPA 180
Db 612 GEPKAGRGVGPVGPVGPAGVCPAGKAGGAGAGQAPGPPGAPGAGRGGQAGSPGQGLPGPP 671
QY 181 GPPGAGRGVGPVGPVGPAGVCPAGKAGGAGAGQAPGPPGAPGAGRGGQAGSPGQGLPGPP 208
Db 672 GPPGAGRGVGPVGPVGPAGVCPAGKAGGAGAGQAPGPPGAPGAGRGGQAGSPGQGLPGPP 699

RESULT 11
Q14046
ID Q14046 PRELIMINARY; PRT; 1160 AA.
AC Q14046
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE COL2A1 protein precursor (Fragment).
GN COL2A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN 1
RP SEQUENCE FROM N.A.
RX MEDLINE=90026318; PubMed=2803268;
RC TISSUE=Cartilage;
RA Baldwin C.T., Reginato A.M., Smith C., Jimenez S.A., Prockop D.J.;
RT "Structure of cDNA clones coding for human type II procollagen. The
RT alpha 1(I) chain is more similar to the alpha 1(I) chain than two
RT other alpha chains of fibrillar collagens."
RL Biochem. J. 262:521-528(1989).
DR EMBL; X16711; CAA34683.1; -.
DR InterPro; IPR008161; C1g_helix.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; C1g_helix; 4.
KW Collagen; Signal.
FT SIGNAL
FT CHAIN
FT NON TER 1160 1160
SQ SEQUENCE 1160 AA; 105630 MW; A7F0523B856C8639 CRC64;

Query Match 72.6%; Score 859; DB 4; Length 1160;
Best Local Similarity 74.0%; Pred. No. 7.7e-52;
Matches 154; Conservative 10; Mismatches 44; Indels 0; Gaps 0;

QY 1 GPPGEPGTGLPGRGGRGSGRFGPCADGVAGPKGAPGERGSPGAPGKSPGAGRP 60
Db 423 GARGEPGVGFIPIGPPGERGAGFNGRFGQDGLAGPKGAPGERGSGLAGPKGANGDPRP 482
QY 61 GEAGLPGAGKLTGSPGSPGDKTGPAGQDGRPPGPPGARGQAGVGMFPKGA 120
Db 483 GEPGLPGAGLTGTPDAGPGKGVGSPGAPCEDGRPPGPGQARGQGVGMFPKGAN 542
QY 121 GEPKAGRGVGPVGPVGPAGVCPAGKAGGAGAGQAPGPPGAPGAGRGGQAGSPGQGLPGPA 180
Db 543 GEPKAGRGVGPVGPVGPAGVCPAGKAGGAGAGQAPGPPGAPGAGRGGQAGSPGQGLPGPP 602
QY 181 GPPGAGRGVGPVGPVGPAGVCPAGKAGGAGAGQAPGPPGAPGAGRGGQAGSPGQGLPGPP 208
Db 603 GPPGAGRGVGPVGPVGPAGVCPAGKAGGAGAGQAPGPPGAPGAGRGGQAGSPGQGLPGPP 699

RESULT 12
Q14046
ID Q14046 PRELIMINARY; PRT; 1269 AA.
AC Q14046
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE COL2A1 protein precursor (Fragment).
GN COL2A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN 1
RP SEQUENCE FROM N.A.
RX MEDLINE=90026318; PubMed=2803268;
RC TISSUE=Cartilage;
RA Baldwin C.T., Reginato A.M., Smith C., Jimenez S.A., Prockop D.J.;
RT "Structure of cDNA clones coding for human type II procollagen. The
RT alpha 1(I) chain is more similar to the alpha 1(I) chain than two
RT other alpha chains of fibrillar collagens."
RL Biochem. J. 262:521-528(1989).
DR EMBL; X16711; CAA34683.1; -.
DR InterPro; IPR008161; C1g_helix.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; C1g_helix; 4.
KW Collagen; Signal.
FT SIGNAL
FT CHAIN
FT NON TER 1269 1269
SQ SEQUENCE 1269 AA; 120116 MW; 446EF91BEB8BF6A CRC64;

Query Match 72.6%; Score 859; DB 13; Length 1269;
Best Local Similarity 73.6%; Pred. No. 8.4e-52;
Matches 153; Conservative 11; Mismatches 44; Indels 0; Gaps 0;

QY 1 GPPGEPGTGLPGRGGRGSGRFGPCADGVAGPKGAPGERGSPGAPGKSPGAGRP 60
Db 274 GARGEPGAGVGPVGPGRGAGFNGRFGQDGLAGPKGAPGERGSGLAGPKGANGDPRP 333
QY 61 GEAGLPGAGKLTGSPGSPGDKTGPAGQDGRPPGPPGARGQAGVGMFPKGA 120
Db 334 GEPGLPGAGLTGTPDAGPGKGVGSPGAPCEDGRPPGPGQARGQGVGMFPKGAN 393
QY 121 GEPKAGRGVGPVGPVGPAGVCPAGKAGGAGAGQAPGPPGAPGAGRGGQAGSPGQGLPGPA 180
Db 394 GEPKAGRGVGPVGPVGPAGVCPAGKAGGAGAGQAPGPPGAPGAGRGGQAGSPGQGLPGPP 453
QY 181 GPPGAGRGVGPVGPVGPAGVCPAGKAGGAGAGQAPGPPGAPGAGRGGQAGSPGQGLPGPP 208
Db 454 GPPGAGRGVGPVGPVGPAGVCPAGKAGGAGAGQAPGPPGAPGAGRGGQAGSPGQGLPGPP 481

RESULT 13
Q90W37
ID Q90W37 PRELIMINARY; PRT; 1420 AA.
AC Q90W37
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Alpha 1 type IIA collagen precursor.
GN COL2A1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN 1
RP SEQUENCE FROM N.A.
RX TISSUE=Sternal;
RA Cai Xia X., Yongzhi X., Siqi G., Yiyang S.;
RT Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; AY046949; AAK38621.1; -.
DR EMBL; AY046949; AAK38621.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002181; Fibrinogen_C.
DR InterPro; IPR000885; Fib_collagen_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; C1g_helix; 6.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
KW Collagen; Signal.

```

```
FT SIGNAL 1 25 POTENTIAL.
SQ SEQUENCE 1420 AA; 134999 MW; 88D9AAB17F214FF5 CRC64;

Query Match 72.6%; Score 859; DB 13; Length 1420;
Best Local Similarity 73.6%; Pred. No. 9.3e-52;
Matches 153; Conservative 11; Mismatches 44; Indels 0; Gaps 0;

QY 1 GPPGEPGPTGLPGPPGGRGSGRFPAGDGVAGPKGPGACRGSPGPGAGKSGPGEAGRP 60
Db 425 GARGEPGAAGVPVPPGGERGAPNGRFPQDGLAGPKGAPGERGPAGLAGPKGATGDPGRP 484

QY 61 GEAGLPGAKGLTSGSPGPGKTPGPGPAGODGRPGPPGPGAGQAGVMPGPKGAA 120
Db 492 GARGEPGVGPIPGPGERGAPGNRFPFGDGLAGPKGAPGERGPGSLAGPKGANGPGRP 551

QY 121 GEPKAGRGVPGPPGAVGPGAGKDGAGAGQAGPPGAGPAGERGEOGAGSPGFGQGLPGPA 180
Db 612 GEPKAGRGVPGPPGAVGPGAGKDGAGAGQAGPPGAGPAGERGEOGAGSPGFGQGLPGPA 180

QY 181 GPPGEGAGKPGEGQVPGDGLGAPGSPGAG 208
Db 672 GPPGEGAGKPGEGQVPGDGLGAPGSPGAG 208

RESULT 15
Q91717 PRELIMINARY; PRT; 1486 AA.
AC Q91717
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)

SEQUENCE FROM N.A.
TISSUE=Blood;
MEDLINE=91153296; PubMed=1999183;
Huang M.C., Seyer J.M., Thompson J.P., Spinella D.G., Cheah K.S.,
Kang A.H.;
"Genomic organization of the human procollagen alpha 1(II) collagen
gene.";
Eur. J. Biochem. 195:593-600(1991).
[6]
SEQUENCE FROM N.A.
TISSUE=Blood;
MEDLINE=92344585;
Vikkula M., Metsaranta M., Syvanen A.C., Ala-Kokko L., Vuorio E.,
Peltonen L.;
"Structural analysis of the regulatory elements of the type-II
procollagen gene. Conservation of promoter and first intron sequences
between human and mouse.";
Biochem. J. 285:0-0(0).
[7]
SEQUENCE FROM N.A.
TISSUE=Blood;
MEDLINE=97104294; PubMed=8948452;
Ala-Kokko L., Kvist A.P., Metsaranta M., Kivirikko K.I.,
de Crombrughe B., Prockop D.J., Vuorio E.;
"Conservation of the sizes of 53 introns and over 100 intronic
sequences for the binding of common transcription factors in the human
and mouse genes for type II procollagen (COL2A1).";
Biochem. J. 308:0-0(0).
ENBL; LI0347; AAC41772.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib collagen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF00093; vwc; 1.
DR ProDom; PD000007; Clg helix; 4.
DR ProDom; PD002078; Fib collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; WFC1; 1.
DR PROSITE; PS0184; WFC2; 1.
KW Collagen.
SQ SEQUENCE 1487 AA; 141772 MW; 0B7E79D46BDAFA97 CRC64;

Query Match 72.6%; Score 859; DB 4; Length 1487;
Best Local Similarity 74.0%; Pred. No. 9.6e-52;
Matches 154; Conservative 10; Mismatches 44; Indels 0; Gaps 0;

QY 1 GPPGEPGPTGLPGPPGGRGSGRFPAGDGVAGPKGPGACRGSPGPGAGKSGPGEAGRP 60
Db 492 GARGEPGVGPIPGPGERGAPGNRFPFGDGLAGPKGAPGERGPGSLAGPKGANGPGRP 551

QY 61 GEAGLPGAKGLTSGSPGPGKTPGPGPAGODGRPGPPGPGAGQAGVMPGPKGAA 120
Db 552 GEPGLPARGLTGTRPDAGPQGVGSGAPGDPGRPPGPGQAGQGVMPGPGPKGAN 611

QY 121 GEPKAGRGVPGPPGAVGPGAGKDGAGAGQAGPPGAGPAGERGEOGAGSPGFGQGLPGPA 180
Db 612 GEPKAGRGVPGPPGAVGPGAGKDGAGAGQAGPPGAGPAGERGEOGAGSPGFGQGLPGPP 671

QY 181 GPPGEGAGKPGEGQVPGDGLGAPGSPGAG 208
Db 672 GPPGEGAGKPGEGQVPGDGLGAPGSPGAG 208

RESULT 15
Q91717 PRELIMINARY; PRT; 1486 AA.
AC Q91717
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)

SEQUENCE FROM N.A.
TISSUE=Blood;
MEDLINE=90026318; PubMed=2803268;
Baldwin C.T., Reginato A.M., Smith C., Jimenez S.A., Prockop D.J.;
"Structure of cDNA clones coding for human type II procollagen. The
alpha 1(II) chain is more similar to the alpha 1(I) chain than two
other alpha chains of fibrillar collagens.";
Biochem. J. 262:521-528(1989).
[3]
SEQUENCE FROM N.A.
TISSUE=Blood;
MEDLINE=89325561; PubMed=2753125;
Vikkula M., Peltonen L.;
"Structural analyses of the polymorphic area in type II collagen
gene.";
FEBS Lett. 250:171-174(1989).
[4]
SEQUENCE FROM N.A.
TISSUE=Blood;
MEDLINE=91184811; PubMed=2081599;
Ryan M.C., Sieraki M., Sandell L.J.;
"The human type II procollagen gene: identification of an additional
protein-coding domain and location of potential regulatory sequences
in the promoter and first intron.";
Genomics 8:41-48(1990).
[5]
SEQUENCE FROM N.A.
```

DT	01-OCT-2003 (TreeBLrel. 25, Last annotation update)
DE	Alpha-1 type II collagen.
DN	COL2A1.
GN	Xenopus laevis (African clawed frog).
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC	Xenopodidae; Xenopus.
OX	NCBI_TaxID=8355;
LN	[1]
RP	SEQUENCE FROM N.A.
RA	MEDLINE=92011898; PubMed=191853;
RX	Su M.W., Suzuki H.R., Bieker J.J., Soltursh M., Ramirez F.;
RT	"Expression of two nonallelic type II procollagen genes during Xenopus
RT	laevis embryogenesis is characterized by stage-specific production of
RT	alternatively spliced transcripts.";
RL	J. Cell Biol. 115:585-575(1991).
DR	EMBL; M63595; AAA49678.1; -.
DR	PIR; A40333; A40333.
DR	PIR; B40333; B40333.
DR	GO; GO:0005581; C:collagen; IEA.
DR	GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR	InterPro; IPR008161; C1q helix.
DR	InterPro; IPR008160; Collagen.
DR	InterPro; IPR000885; Fib_collagen_C.
DR	InterPro; IPR001007; VWFC.
DR	Pfam; PF01410; COLFI; 1.
DR	Pfam; PF01391; Collagen; 18.
DR	Pfam; PF00093; vwc; 1.
DR	ProDom; PD000007; C1q helix; 4.
DR	ProDom; PD002078; Fib_collagen_C; 1.
DR	SMART; SM00038; COLFI; 1.
DR	SMART; SM00214; VWC; 1.
DR	PROSITE; PS01208; VWFC.1; 1.
DR	PROSITE; PSS0184; VWFC_2; 1.
KW	Collagen.
SQ	SEQUENCE 1486 AA; 142263 MW; 4A4A95772341042F CRC64;
Query Match 71.8%; Score 850; DB 13; Length 1486;	
Best Local Similarity 73.1%; Pred.No.4e-51;	
Matches 152; Conservative 11; Mismatches 45; Indels 0; Gaps 0;	
QY	1 GPPGPGGTGLPGPPGGRGPGSGRPFPGADVGAGKXPAGERGSPGAPKSGPGEAGRP 60
Db	494 GARGEPGAAGPNGPPGERGAPENRGPPGQDLGAGKXGAPGERGVPLCGPKGNGDPGR 553
QY	61 GEAGLPGAKGLTGSFSGPPDGKTTGPPGPAQDGRGPPGPPGARGAQAGVMGFPGPGAA 120
Db	554 GEPGLPGARGUTGRPGDAGPGQKVGPSGASGEDRGPPGPGQARGQGQVWGFPGPKGAN 613
QY	121 GEPGKAGRGVGPFPAGVAGPKDGEAGAGCPGPPAGPAGERGEOGPAGSGPQGLPGPA 180
Db	614 GEPGKAGEKGLVAFGLRGFLPKDGEVSGFNGPAGPAGERGEQGPFGPSGFQGLPGPP 673
QY	181 GPPGEGAKPGQGVPGDLGAPGSPGPAG 208
b	674 GSPGEGGKPGDQGVPGEGAGFLVGPGRG 701

Search completed: September 24, 2004, 11:11:12
Job time : 24.0943 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2004, 11:07:31 ; Search time 25,8127 Seconds
(without alignments)
2503.588 Million cell updates/sec

Title: US-10-658-989a-1

Perfect score: 1184

Sequence: 1 GPPGEPPTGLPQPPGRRG.....GEQGVGDLGAPGSPGAGG 209

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1349238 seqs, 32158718 residues

Total number of hits satisfying chosen parameters: 1349238

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/1/pubaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/1/pubaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/1/pubaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/1/pubaa/US09C_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/1/pubaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/1/pubaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/1/pubaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/1/pubaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/1/pubaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/1/pubaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1149	97.0	501	12	US-10-232-175-17
2	1149	97.0	1057	16	US-10-104-889-16
3	1149	97.0	1057	16	US-10-104-889-20
4	1149	97.0	1107	16	US-10-104-889-11
5	1149	97.0	1169	16	US-10-104-889-6
6	1149	97.0	1171	16	US-10-104-889-8
7	1149	97.0	1341	14	US-10-058-224-18
8	1149	97.0	1388	16	US-10-104-889-10
9	1149	97.0	1461	16	US-10-468-091-25
10	1149	97.0	1464	12	US-09-918-715-261
11	1149	97.0	1464	14	US-10-060-036-159
12	1149	97.0	1464	14	US-10-171-311-36
13	1149	97.0	1464	14	US-10-216-705-21
14	1149	97.0	1464	14	US-10-149-352-2
15	1149	97.0	1464	14	US-10-177-293-65

16	1149	97.0	1464	14	US-10-301-822-28	Sequence 28, Appl
17	1149	97.0	1464	15	US-10-291-265-243	Sequence 243, Appl
18	1149	97.0	1464	16	US-10-357-851-1	Sequence 1, Appl
19	1149	97.0	1464	16	US-10-358-024-1	Sequence 1, Appl
20	1149	97.0	1464	16	US-10-734-564-79	Sequence 79, Appl
21	1144	96.6	1449	15	US-10-402-089-8	Sequence 8, Appl
22	1144	96.6	1449	15	US-10-402-072A-8	Sequence 8, Appl
23	1141	96.4	1463	15	US-10-402-089-2	Sequence 2, Appl
24	1141	96.4	1463	15	US-10-402-072A-2	Sequence 2, Appl
25	1116	94.3	595	15	US-10-342-331-48	Sequence 48, Appl
26	1116	94.3	595	15	US-10-342-331-50	Sequence 50, Appl
27	1116	94.3	822	15	US-10-342-331-49	Sequence 49, Appl
28	1116	94.3	822	15	US-10-468-091-26	Sequence 26, Appl
29	908	76.7	492	16	US-10-639-286-12	Sequence 12, Appl
30	866	73.1	1017	16	US-10-639-286-10	Sequence 10, Appl
31	862	72.8	1418	14	US-10-058-124-20	Sequence 20, Appl
32	862	72.8	1418	16	US-10-468-091-5	Sequence 5, Appl
33	851	71.9	1014	14	US-10-194-441A-1	Sequence 1, Appl
34	841.5	71.1	714	14	US-10-233-885-44	Sequence 44, Appl
35	841.5	71.1	714	14	US-10-231-581-44	Sequence 44, Appl
36	841.5	71.1	714	15	US-10-326-508A-15	Sequence 15, Appl
37	839	70.9	1014	14	US-10-194-441A-48	Sequence 48, Appl
38	839	70.9	1487	16	US-10-468-091-6	Sequence 6, Appl
39	814	68.8	185	12	US-10-232-175-20	Sequence 20, Appl
40	814	68.8	251	12	US-10-232-175-21	Sequence 21, Appl
41	814	68.8	500	12	US-10-232-175-22	Sequence 22, Appl
42	814	68.8	662	12	US-10-232-175-33	Sequence 33, Appl
43	787.5	66.5	1466	15	US-10-402-089-12	Sequence 12, Appl
44	787.5	66.5	1466	15	US-10-402-072A-12	Sequence 12, Appl
45	786	66.4	1496	14	US-10-177-293-70	Sequence 70, Appl

ALIGNMENTS

RESULT 1
US-10-232-175-17
; Sequence 17, Application US/10232175
; Publication No. US20030064074A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Kivirikko, Kari I.
; APPLICANT: Neff, Thomas B.
; APPLICANT: Olsen, David R.
; APPLICANT: Polarek, James W.
; TITLE OF INVENTION: RECOMBINANT GELATINS IN VACCINES
; FILE REFERENCE: FG0224
; CURRENT APPLICATION NUMBER: US/10/232,175
; PRIOR FILING DATE: 2002-08-30
; CURRENT FILING DATE: US/09/710,249
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 17
; LENGTH: 501
; TYPE: PRT
; ORGANISM: human
US-10-232-175-17

Query Match	97.0%	Score 1149;	DB 12;	Length 501;
Best Local Similarity	93.6%	Pred. No. 7.5e-59;		
Matches 206;	Conservative 0;	Mismatches 2;	Indels 12;	Gaps 1;
Qy	1	GPP-----GEPGPTGLPQPPGRRGQPPGFCADCVAGPKGPKAGRSGPSPA 48		
Db	280	GPPGAGEGKRCARSEPFTGLPQPPGRRGQPPGFCADCVAGPKGPKAGRSGPSPA 339		
Qy	49	GPXGSGEAGRPCEAGLPKAGLTGSPGSPGDKTGTGPPACQDGRPPGPPGARGOA 108		
Db	340	GPXGSGEAGRPCEAGLPKAGLTGSPGSPGDKTGTGPPACQDGRPPGPPGARGOA 399		
Qy	109	GVWGFPGPKAGCEPKAGRGVPPGAVGAGKDGACAGCPGPGPAGPAGRGFGGPA 168		

Db 400 GWMGPPGKAGGPKAGRGVPGPPGAVGAGPKAGDGEAGAGQGGPPGAGPAGERGEGCPA 459

QY 169 GSPGQGLPGPAGPGEAGKPGEGQVPGDGLGAPGSPGAG 208
Db 460 GSPGQGLPGPAGPGEAGKPGEGQVPGDGLGAPGSPGAG 499

RESULT 2

US-10-104-889-16
; Sequence 16, Application US/10104889
; Publication No. US20040086961A1
; GENERAL INFORMATION:

APPLICANT: GRUSKIN, ELLIOT A.
BROKAW, JANE
ZHANG, GUANGHUI
PAOLELLA, DAVID

TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:

ADDRESSEE: DILWORTH & BARRESE
STREET: 333 EARLE OVINGTON BOULEVARD
CITY: UNIONDALE

STATE: NY
COUNTRY: U.S.A.
ZIP: 11553

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/104,889
FILING DATE: 22-Mar-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/169,768
FILING DATE: 09-OCT-1998
ATTORNEY/AGENT INFORMATION:

NAME: STEEN, JEFFREY S
TELECOMMUNICATION INFORMATION:

TELEPHONE: (516) 228-8484
TELEFAX: (516) 228-8516

INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:
LENGTH: 1057 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 16:

US-10-104-889-16

Query Match 97.0%; Score 1149; DB 16; Length 1057;
Best Local Similarity 93.6%; Pred. No. 1.4e-58;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPGPPGGRGPPGADGVAGPKGPPGAGERSGCPA 48

Db 297 GPPGAGEGKRGAGEGPGTGLPGPPGGRGPPGADGVAGPKGPPGAGERSGCPA 356

QY 49 GPKSGPAGRPAGLPGAKGLTGSPGSPGDPGKTGPPAGDGRPPGPPGARGQA 108

Db 357 GPKSGPAGRPAGLPGAKGLTGSPGSPGDPGKTGPPAGDGRPPGPPGARGQA 416

QY 109 GVMGPPGKGAAGBPGKAGRGVPPGAVGPAKDGAGAGQGGPPGAGERSGCPA 168

Db 417 GVMGPPGKGAAGBPGKAGRGVPPGAVGPAKDGAGAGQGGPPGAGERSGCPA 476

QY 169 GSPGQGLPGPAGPGEAGKPGEGQVPGDGLGAPGSPGAG 208

Db 477 GSPGQGLPGPAGPGEAGKPGEGQVPGDGLGAPGSPGAG 516

RESULT 3

US-10-104-889-20
; Sequence 20, Application US/10104889
; Publication No. US20040086961A1
; GENERAL INFORMATION:

APPLICANT: GRUSKIN, ELLIOT A.
BROKAW, JANE
ZHANG, GUANGHUI
PAOLELLA, DAVID

TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:

ADDRESSEE: DILWORTH & BARRESE
STREET: 333 EARLE OVINGTON BOULEVARD
CITY: UNIONDALE

STATE: NY
COUNTRY: U.S.A.
ZIP: 11553

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/104,889
FILING DATE: 22-Mar-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/169,768
FILING DATE: 09-OCT-1998
ATTORNEY/AGENT INFORMATION:

NAME: STEEN, JEFFREY S
TELECOMMUNICATION INFORMATION:

TELEPHONE: (516) 228-8484
TELEFAX: (516) 228-8516

INFORMATION FOR SEQ ID NO: 20:

SEQUENCE CHARACTERISTICS:
LENGTH: 1057 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 20:

US-10-104-889-20

Query Match 97.0%; Score 1149; DB 16; Length 1057;
Best Local Similarity 93.6%; Pred. No. 1.4e-58;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPGPPGGRGPPGADGVAGPKGPPGAGERSGCPA 48

Db 297 GPPGAGEGKRGAGEGPGTGLPGPPGGRGPPGADGVAGPKGPPGAGERSGCPA 356

QY 49 GPKSGPAGRPAGLPGAKGLTGSPGSPGDPGKTGPPAGDGRPPGPPGARGQA 108

Db 357 GPKSGPAGRPAGLPGAKGLTGSPGSPGDPGKTGPPAGDGRPPGPPGARGQA 416

QY 109 GVMGPPGKGAAGBPGKAGRGVPPGAVGPAKDGAGAGQGGPPGAGERSGCPA 168

Db 417 GVMGPPGKGAAGBPGKAGRGVPPGAVGPAKDGAGAGQGGPPGAGERSGCPA 476

QY 169 GSPGQGLPGPAGPGEAGKPGEGQVPGDGLGAPGSPGAG 208

Db 477 GSPGQGLPGPAGPGEAGKPGEGQVPGDGLGAPGSPGAG 516

RESULT 4

US-10-104-889-11
; Sequence 11, Application US/10104889
; Publication No. US20040086961A1
; GENERAL INFORMATION:

GENERAL INFORMATION:
APPLICANT: GRUSKIN, ELLIOT A.
BUECHTER, DOUGLAS
BROKAW, JANE
ZHANG, GUANGHUI
FLORELLA, DAVID
TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESS: DILWORTH & BARRESE
STREET: 333 EARLE OVINGTON BOULEVARD
CITY: UNIONDALE
STATE: NY

COUNTRY: U.S.A.
ZIP: 11553
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/104,889
FILING DATE: 22-Mar-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/169,768
FILING DATE: 09-OCT-1998
ATTORNEY/AGENT INFORMATION:
NAME: STEEN, JEFFREY S
TELEPHONE: (516) 228-8484
TELEFAX: (516) 228-8516
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1171 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-10-104-889-8

Query Match 97.0%; Score 1149; DB 16; Length 1171;
Best Local Similarity 93.6%; Pred. No. 1.5e-56;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
QY 1 GPP-----GEPGTGLRPPGGERGSGRFPFGADGVAGPKGPGAGERSGPGA 48
Db 297 GPPGAGEGKKGARGEPGTGLRPPGGERGSGRFPFGADGVAGPKGPGAGERSGPGA 356
QY 49 GPKSGPGEAGRGEAGLPCAKGLTSGSPGPDGKTGPPGAGDGRPPGPPGARGQA 108
Db 357 GPKSGPGEAGRGEAGLPCAKGLTSGSPGPDGKTGPPGAGDGRPPGPPGARGQA 416
QY 109 GVMFPFGKGAAGPGKAGRGVPPGAVGPAKGDEAGAGQPPGAGPAGERGEGGPA 168
Db 417 GVMFPFGKGAAGPGKAGRGVPPGAVGPAKGDEAGAGQPPGAGPAGERGEGGPA 476
QY 169 GSPFGQGLPGPAGPPGEGAGKFGEGVPGDGLGAPGSPGAG 208
Db 477 GSPFGQGLPGPAGPPGEGAGKFGEGVPGDGLGAPGSPGARG 516

RESULT 7
US-10-058-124-18
Sequence 18, Application US/10058124
Publication No. US20030119058A1
GENERAL INFORMATION:
APPLICANT: Qvist, Per
Bonde, Martin
TITLE OF INVENTION: A Method for Assaying Collagen Fragments
in Body Fluids, A Test Kit and Means for Carrying Out the
Method and Use of the Method to Diagnose the Presence of
Disorders Associated with the Metabolism of
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby PC
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/058,124
FILING DATE: 29-Jan-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/570,573
FILING DATE: 2002-MAY-12
APPLICATION NUMBER: 08/187,319
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Gogoris, Adda C
REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 4305/08701
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 1341 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: COLLAGEN ALPHA 1 (I)
SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-10-058-124-18
Query Match 97.0%; Score 1149; DB 14; Length 1341;
Best Local Similarity 93.6%; Pred. No. 1.7e-58;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
QY 1 GPP-----GEPGTGLRPPGGERGSGRFPFGADGVAGPKGPGAGERSGPGA 48
Db 334 GPPGAGEGKKGARGEPGTGLRPPGGERGSGRFPFGADGVAGPKGPGAGERSGPGA 393
QY 49 GPKSGPGEAGRGEAGLPCAKGLTSGSPGPDGKTGPPGAGDGRPPGPPGARGQA 108
Db 394 GPKSGPGEAGRGEAGLPCAKGLTSGSPGPDGKTGPPGAGDGRPPGPPGARGQA 453
QY 109 GVMFPFGKGAAGPGKAGRGVPPGAVGPAKGDEAGAGQPPGAGPAGERGEGGPA 168
Db 454 GVMFPFGKGAAGPGKAGRGVPPGAVGPAKGDEAGAGQPPGAGPAGERGEGGPA 513
QY 169 GSPFGQGLPGPAGPPGEGAGKFGEGVPGDGLGAPGSPGAG 208
Db 514 GSPFGQGLPGPAGPPGEGAGKFGEGVPGDGLGAPGSPGARG 553

RESULT 8
US-10-104-889-10
Sequence 10, Application US/10104889
Publication No. US20040086961A1
GENERAL INFORMATION:
APPLICANT: GRUSKIN, ELLIOT A.
BUECHTER, DOUGLAS
BROKAW, JANE
ZHANG, GUANGHUI
PAOLELLA, DAVID
TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: DILWORTH & BARRESE
STREET: 333 EARLE OVINGTON BOULEVARD
CITY: NY
STATE: NY
COUNTRY: U.S.A.
ZIP: 11553
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 22-Mar-2002
CLASSIFICATION: <Unknown>
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/09/169,768
FILING DATE: 09-OCT-1998
ATTORNEY/AGENT INFORMATION:
NAME: STEEN, JEFFREY S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 228-8484
TELEFAX: (516) 228-8516
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1388 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-10-104-889-10:
Query Match 97.0%; Score 1149; DB 16; Length 1388;
Best Local Similarity 93.6%; Pred. No. 1.7e-58;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
QY 1 GPP-----GEPGPTGLPGERGGRGSGRFGADGVAGPKGAGERGSPGPA 48
DB 297 GPPGAGEGKRGARGEPGPTGLPGERGGRGSGRFGADGVAGPKGAGERGSPGPA 356
QY 49 GPKGSPGAGRPGEAGLPGAKGLTSGPSGPDGKTGTPGAGQDGRPPGPPGARGQA 108
DB 357 GPKGSPGAGRPGEAGLPGAKGLTSGPSGPDGKTGTPGAGQDGRPPGPPGARGQA 416
QY 109 GVMGPPGKGAAGEPKAGRGVPPGAVGPKGDKGNGAGQGGPPGAGPAGERGEGGPA 168
DB 417 GVMGPPGKGAAGEPKAGRGVPPGAVGPKGDKGNGAGQGGPPGAGPAGERGEGGPA 476
QY 169 GSPGFQGLPGPAGPGEAGKPGEGQGVPGDLGAPGSPGAG 208
DB 477 GSPGFQGLPGPAGPGEAGKPGEGQGVPGDLGAPGSPGARG 516
RESULT 9
US-10-468-091-25
Sequence 25, Application US/10468091
Publication No. US20040157329A1
GENERAL INFORMATION:
APPLICANT: ADP Pharmaceutical Pty Limited
TITLE OF INVENTION: Matrix gene expression in chondrogenesis
FILE REFERENCE: 500311
CURRENT APPLICATION NUMBER: US/10/468,091
CURRENT FILING DATE: 2003-08-13
PRIOR APPLICATION NUMBER: AU PR3116
PRIOR FILING DATE: 2001-02-15
NUMBER OF SEQ ID NOS: 42
SOFTWARE: Patent in version 3.1
SEQ ID NO 25
LENGTH: 1461
TYPE: PRT
ORGANISM: Homo sapiens
US-10-468-091-25
Query Match 97.0%; Score 1149; DB 16; Length 1461;
Best Local Similarity 93.6%; Pred. No. 1.8e-58;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
QY 1 GPP-----GEPGPTGLPGERGGRGSGRFGADGVAGPKGAGERGSPGPA 48

DB 455 GPPGAGEGKRGARGEPGPTGLPGERGGRGSGRFGADGVAGPKGAGERGSPGPA 514
QY 49 GPKGSPGAGRPGEAGLPGAKGLTSGPSGPDGKTGTPGAGQDGRPPGPPGARGQA 108
DB 515 GPKGSPGAGRPGEAGLPGAKGLTSGPSGPDGKTGTPGAGQDGRPPGPPGARGQA 574
QY 109 GVMGPPGKGAAGEPKAGRGVPPGAVGPKGDKGNGAGQGGPPGAGPAGERGEGGPA 168
DB 575 GVMGPPGKGAAGEPKAGRGVPPGAVGPKGDKGNGAGQGGPPGAGPAGERGEGGPA 634
QY 169 GSPGFQGLPGPAGPGEAGKPGEGQGVPGDLGAPGSPGAG 208
DB 635 GSPGFQGLPGPAGPGEAGKPGEGQGVPGDLGAPGSPGARG 674
RESULT 10
US-09-918-715-261
Sequence 261, Application US/09918715
Publication No. US20030017157A1
GENERAL INFORMATION:
APPLICANT: Brad St. Croix
APPLICANT: Bert Vogelstein
APPLICANT: Kenneth Kinzler
TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS
FILE REFERENCE: 1107, 00134
CURRENT APPLICATION NUMBER: US/09/918,715
CURRENT FILING DATE: 2001-08-01
PRIOR APPLICATION NUMBER: 60/222,599
PRIOR FILING DATE: 2000-08-02
PRIOR APPLICATION NUMBER: 60/224,360
PRIOR FILING DATE: 2000-08-11
PRIOR APPLICATION NUMBER: 60/282,850
PRIOR FILING DATE: 2000-04-11
NUMBER OF SEQ ID NOS: 358
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 261
LENGTH: 1464
TYPE: PRT
ORGANISM: Homo sapiens
US-09-918-715-261
Query Match 97.0%; Score 1149; DB 12; Length 1464;
Best Local Similarity 93.6%; Pred. No. 1.8e-58;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
QY 1 GPP-----GEPGPTGLPGERGGRGSGRFGADGVAGPKGAGERGSPGPA 48
DB 458 GPPGAGEGKRGARGEPGPTGLPGERGGRGSGRFGADGVAGPKGAGERGSPGPA 517
QY 49 GPKGSPGAGRPGEAGLPGAKGLTSGPSGPDGKTGTPGAGQDGRPPGPPGARGQA 108
DB 518 GPKGSPGAGRPGEAGLPGAKGLTSGPSGPDGKTGTPGAGQDGRPPGPPGARGQA 577
QY 109 GVMGPPGKGAAGEPKAGRGVPPGAVGPKGDKGNGAGQGGPPGAGPAGERGEGGPA 168
DB 578 GVMGPPGKGAAGEPKAGRGVPPGAVGPKGDKGNGAGQGGPPGAGPAGERGEGGPA 637
QY 169 GSPGFQGLPGPAGPGEAGKPGEGQGVPGDLGAPGSPGAG 208
DB 638 GSPGFQGLPGPAGPGEAGKPGEGQGVPGDLGAPGSPGARG 677
RESULT 11
US-10-060-036-159
Sequence 159, Application US/10060036
Publication No. US20030073144A1
GENERAL INFORMATION:
APPLICANT: Benson, Darin R.
APPLICANT: Kalos, Michael D.
APPLICANT: Lodes, Michael J.
APPLICANT: Persing, David H.
APPLICANT: Hepler, William T.
APPLICANT: Jiang, Yuqiu

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 21021.566
; CURRENT APPLICATION NUMBER: US/10/060,036
; CURRENT FILING DATE: 2002-01-30
; NUMBER OF SEQ ID NOS: 4560
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 159
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-060-036-159

Query Match 97.0%; Score 1149; DB 14; Length 1464;
Best Local Similarity 93.6%; Pred. No. 1.8e-58;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEPGTLGPPGGERGPGSRGPPGADGVAGPKGAGERGSPGPA 48
DB 458 GPPGAGEGKRGAGEGPGTGLPGGGERGPGSRGPPGADGVAGPKGAGERGSPGPA 517
QY 49 GPKSGPGEAGRPGEAGLPGAKGLTGSPGSPGDPGKTGPPGAGQDGRPPGPPGARGQA 108
DB 518 GPKSGPGEAGRPGEAGLPGAKGLTGSPGSPGDPGKTGPPGAGQDGRPPGPPGARGQA 577
QY 109 GVMGPPGPKGAAGEPKKAGERGVPPGCAVGPAGKDGAGAGAGQGGPPGAPGAGERGEOGPA 168
DB 578 GVMGPPGPKGAAGEPKKAGERGVPPGCAVGPAGKDGAGAGAGQGGPPGAPGAGERGEOGPA 637
QY 169 GSPGFQGLPGPAGPGEAGKPGEGQGVPGDLGAPGSPGAG 208
DB 638 GSPGFQGLPGPAGPGEAGKPGEGQGVPGDLGAPGSPGAG 677

RESULT 12

US-10-171-311-36

; Sequence 36, Application US/10171311
; Publication No. US20030087270A1

; GENERAL INFORMATION:

; APPLICANT: Schlegel, Robert

; APPLICANT: Chen, Yan

; APPLICANT: Zhao, Xumei

; APPLICANT: Monahan, John

; APPLICANT: Kamatkar, Shubhangi

; APPLICANT: Glatt, Karen

; APPLICANT: Gannavarapu, Manjula

; APPLICANT: Hoerish, Sebastian

; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR

; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY

; TITLE OF INVENTION: OF CERVICAL CANCER

; FILE REFERENCE: MRI-035

; CURRENT APPLICATION NUMBER: US/10/171.311

; CURRENT FILING DATE: 2002-06-12

; PRIOR APPLICATION NUMBER: US 60/298,159

; PRIOR FILING DATE: 2001-06-13

; PRIOR APPLICATION NUMBER: US 60/298,155

; PRIOR FILING DATE: 2001-06-13

; PRIOR APPLICATION NUMBER: US 60/335,936

; PRIOR FILING DATE: 2001-11-14

; NUMBER OF SEQ ID NOS: 238

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 36

; LENGTH: 1464

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-171-311-36

Query Match 97.0%; Score 1149; DB 14; Length 1464;
Best Local Similarity 93.6%; Pred. No. 1.8e-58;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEPGTLGPPGGERGPGSRGPPGADGVAGPKGAGERGSPGPA 48

DB 458 GPPGAGEGKRGAGEGPGTGLPGGGERGPGSRGPPGADGVAGPKGAGERGSPGPA 517
QY 49 GPKSGPGEAGRPGEAGLPGAKGLTGSPGSPGDPGKTGPPGAGQDGRPPGPPGARGQA 108
DB 518 GPKSGPGEAGRPGEAGLPGAKGLTGSPGSPGDPGKTGPPGAGQDGRPPGPPGARGQA 577
QY 109 GVMGPPGPKGAAGEPKKAGERGVPPGCAVGPAGKDGAGAGAGQGGPPGAPGAGERGEOGPA 168
DB 578 GVMGPPGPKGAAGEPKKAGERGVPPGCAVGPAGKDGAGAGAGQGGPPGAPGAGERGEOGPA 637
QY 169 GSPGFQGLPGPAGPGEAGKPGEGQGVPGDLGAPGSPGAG 208
DB 638 GSPGFQGLPGPAGPGEAGKPGEGQGVPGDLGAPGSPGAG 677

RESULT 13

US-10-216-705-21

; Sequence 21, Application US/10216705

; Publication No. US20030096973A1

; GENERAL INFORMATION:

; APPLICANT: Meristem Therapeutics, S.A.

; TITLE OF INVENTION: Recombinant Collagens and Derived Proteins Produced by Plants, Met

; FILE REFERENCE: 1149-3 DIV

; CURRENT APPLICATION NUMBER: US/10/216,705

; CURRENT FILING DATE: 2002-08-09

; PRIOR APPLICATION NUMBER: US 09/331,347

; PRIOR FILING DATE: 1999-08-17

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: Patent in version 3.1

; SEQ ID NO 21

; LENGTH: 1464

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-216-705-21

Query Match 97.0%; Score 1149; DB 14; Length 1464;
Best Local Similarity 93.6%; Pred. No. 1.8e-58;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEPGTLGPPGGERGPGSRGPPGADGVAGPKGAGERGSPGPA 48
DB 458 GPPGAGEGKRGAGEGPGTGLPGGGERGPGSRGPPGADGVAGPKGAGERGSPGPA 517
QY 49 GPKSGPGEAGRPGEAGLPGAKGLTGSPGSPGDPGKTGPPGAGQDGRPPGPPGARGQA 108
DB 518 GPKSGPGEAGRPGEAGLPGAKGLTGSPGSPGDPGKTGPPGAGQDGRPPGPPGARGQA 577
QY 109 GVMGPPGPKGAAGEPKKAGERGVPPGCAVGPAGKDGAGAGAGQGGPPGAPGAGERGEOGPA 168
DB 578 GVMGPPGPKGAAGEPKKAGERGVPPGCAVGPAGKDGAGAGAGQGGPPGAPGAGERGEOGPA 637
QY 169 GSPGFQGLPGPAGPGEAGKPGEGQGVPGDLGAPGSPGAG 208
DB 638 GSPGFQGLPGPAGPGEAGKPGEGQGVPGDLGAPGSPGAG 677

RESULT 14

US-10-149-352-2

; Sequence 2, Application US/10149352

; Publication No. US20030105050A1

; GENERAL INFORMATION:

; APPLICANT: Beri, Rajinder

; TITLE OF INVENTION: ANTISENSE OLIGONUCLEOTIDES

; FILE REFERENCE: 06275-254US1

; CURRENT APPLICATION NUMBER: US/10/149,352

; CURRENT FILING DATE: 2002-06-10

; PRIOR APPLICATION NUMBER: PCT/US00/04741

; PRIOR FILING DATE: 2000-12-12

; PRIOR APPLICATION NUMBER: GB 9929487.8

; PRIOR FILING DATE: 1999-12-15

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: Patent in Ver. 4.0

US-10-177-293-65

Query Match 97.0%; Score 1149; DB 14; Length 1464;
Best Local Similarity 93.6%; Pred. No. 1.8e-58;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPQPGRGGPGSRGPPGADGVAGPKGPAGRGSPGPA 48
DB 458 GPPGPAGEEGKRGARGEPGPTGLPQPGRGGPGSRGPPGADGVAGPKGPAGRGSPGPA 517

QY 49 GPKGSPGEAGRPGEAGLPGAKGLTGSPPGPDGKTGPPGAGODGRPPGPPGARGQA 108
DB 518 GPKGSPGEAGRPGEAGLPGAKGLTGSPPGPDGKTGPPGAGODGRPPGPPGARGQA 577

QY 109 GVMGFPKGAAGPPGKAGRGVPGPPCAVGPAGKDGAGAGQPPGPPGAGRGEGGPA 168
DB 578 GVMGFPKGAAGPPGKAGRGVPGPPCAVGPAGKDGAGAGQPPGPPGAGRGEGGPA 637

QY 169 GSPGFOGLPGPAGPPGKAGRGEGVPGDGLGAPGSPGAG 208
DB 638 GSPGFOGLPGPAGPPGKAGRGEGVPGDGLGAPGSPGAG 677

Search completed: September 24, 2004, 11:13:23
Job time : 26.8127 sec

US-10-658-989a-1.rapb

Query Match 97.0%; Score 1149; DB 14; Length 1464;
Best Local Similarity 93.6%; Pred. No. 1.8e-58;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPQPGRGGPGSRGPPGADGVAGPKGPAGRGSPGPA 48
DB 458 GPPGPAGEEGKRGARGEPGPTGLPQPGRGGPGSRGPPGADGVAGPKGPAGRGSPGPA 517

QY 49 GPKGSPGEAGRPGEAGLPGAKGLTGSPPGPDGKTGPPGAGODGRPPGPPGARGQA 108
DB 518 GPKGSPGEAGRPGEAGLPGAKGLTGSPPGPDGKTGPPGAGODGRPPGPPGARGQA 577

QY 109 GVMGFPKGAAGPPGKAGRGVPGPPCAVGPAGKDGAGAGQPPGPPGAGRGEGGPA 168
DB 578 GVMGFPKGAAGPPGKAGRGVPGPPCAVGPAGKDGAGAGQPPGPPGAGRGEGGPA 637

QY 169 GSPGFOGLPGPAGPPGKAGRGEGVPGDGLGAPGSPGAG 208
DB 638 GSPGFOGLPGPAGPPGKAGRGEGVPGDGLGAPGSPGAG 677

RESULT 15

US-10-177-293-65

Sequence 65, Application US/10177293
Publication No. US20030124128A1
GENERAL INFORMATION:
APPLICANT: Lillie, James
APPLICANT: Zhao, Xumei
APPLICANT: Ganavaru, Manjula
APPLICANT: Kamatkar, Shubhangi
APPLICANT: Mertens, Maureen
APPLICANT: Myer, Vic
APPLICANT: Wang, Youzhen
APPLICANT: Xu, Yongyao
APPLICANT: Hoersch, Sebastian
APPLICANT: Monahan, John
APPLICANT: Meyers, Rachel E.
APPLICANT: Bast Jr., Robert C.
APPLICANT: Hortobagyi, Gabriel N.
APPLICANT: Pusztai, Lajos
APPLICANT: Meric, Funda
APPLICANT: Sahin, Aysegul
APPLICANT: Mills, Gordon B.
TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
PREVENTION, AND THERAPY OF BREAST CANCER
FILE REFERENCE: MRI-038
CURRENT APPLICATION NUMBER: US/10/177,293
CURRENT FILING DATE: 2002-06-21
PRIOR APPLICATION NUMBER: US 60/299,887
PRIOR FILING DATE: 2001-06-21
PRIOR APPLICATION NUMBER: US 60/301,572
PRIOR FILING DATE: 2001-06-27
PRIOR APPLICATION NUMBER: US 60/306,501
PRIOR FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: US 60/325,002
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US 60/362,585
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US 60/xxx,xxx
PRIOR FILING DATE: 2002-05-14
NUMBER OF SEQ ID NOS: 506
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 65
LENGTH: 1464
TYPE: PRT
ORGANISM: Homo sapiens

US-10-177-293-65

Query Match 97.0%; Score 1149; DB 14; Length 1464;
Best Local Similarity 93.6%; Pred. No. 1.8e-58;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPQPGRGGPGSRGPPGADGVAGPKGPAGRGSPGPA 48
DB 458 GPPGPAGEEGKRGARGEPGPTGLPQPGRGGPGSRGPPGADGVAGPKGPAGRGSPGPA 517

QY 49 GPKGSPGEAGRPGEAGLPGAKGLTGSPPGPDGKTGPPGAGODGRPPGPPGARGQA 108
DB 518 GPKGSPGEAGRPGEAGLPGAKGLTGSPPGPDGKTGPPGAGODGRPPGPPGARGQA 577

QY 109 GVMGFPKGAAGPPGKAGRGVPGPPCAVGPAGKDGAGAGQPPGPPGAGRGEGGPA 168
DB 578 GVMGFPKGAAGPPGKAGRGVPGPPCAVGPAGKDGAGAGQPPGPPGAGRGEGGPA 637

QY 169 GSPGFOGLPGPAGPPGKAGRGEGVPGDGLGAPGSPGAG 208
DB 638 GSPGFOGLPGPAGPPGKAGRGEGVPGDGLGAPGSPGAG 677

Search completed: September 24, 2004, 11:13:23
Job time : 26.8127 sec

This Page Blank (usp10)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2004, 11:06:55 ; Search time 5.27357 Seconds
(without alignments)
2063.622 Million cell updates/sec

Title: US-10-658-989a-1

Perfect score: 1184

Sequence: 1 GPGEGPTGLPGRGG.....GEGVFDLGAQPGSPAGG 209

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1149	97.0	1460	1	Q8XSJ7 canis fami
2	1149	97.0	1464	1	P02452 homo sapien
3	1118	94.4	671	1	P02454 rattus norv
4	1116	94.3	1453	1	P11087 mus musculu
5	1067	90.1	1453	1	P02457 gallus gall
6	913	77.1	747	1	P02459 bos taurus
7	862	72.8	1418	1	P02458 homo sapien
8	839	70.9	1459	1	P28481 mus musculu
9	824	69.6	779	1	P02453 bos taurus
10	794.5	67.1	1262	1	P12105 gallus gall
11	786	66.4	1496	1	P05997 homo sapien
12	776	65.5	1466	1	P02461 homo sapien
13	764	64.5	1372	1	P02466 rattus norv
14	763	64.4	1464	1	P03121 mus musculu
15	761.5	64.3	1049	1	P04258 bos taurus
16	760	64.2	1372	1	Q01149 mus musculu
17	744	62.8	1364	1	P02465 bos taurus
18	739	62.4	1366	1	P08123 homo sapien
19	732	61.8	1366	1	O46392 canis fami
20	703	59.4	1355	1	O42350 rana catesb
21	693.5	58.6	1356	1	O93484 oncorhynch
22	650.5	54.9	1362	1	P02467 gallus gall
23	629.5	53.2	1027	1	P30754 riftia pach
24	606	51.2	1838	1	P20908 homo sapien
25	590.5	49.9	1736	1	P13942 homo sapien
26	584.5	49.4	1736	1	O64739 mus musculu
27	577	48.7	1806	1	P12107 homo sapien
28	576	48.6	754	1	Q28247 canis fami
29	576	48.6	911	1	Q28083 bos taurus
30	575.5	48.6	3124	1	P13944 gallus gall
31	574.5	48.5	675	1	P32017 gallus gall
32	573.5	48.4	1690	1	P53420 homo sapien
33	572.5	48.4	636	1	P13941 rattus norv

ALIGNMENTS

RESULT 1

ID	CALL_CANFA	STANDARD;	PRT;	1460 AA.
AC	Q9XSJ7;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	Collagen alpha 1(I) chain precursor.			
GN	COL1A1.			
OS	Canis familiaris (Dog).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.			
OX	NCBI_Taxid=9615;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND VARIANT OI ALA-208.			
RC	TISSUE=Skin;			
RX	MEDLINE=21023337; PubMed=11147834;			
RA	Campbell B.G., Wootton J.A.M., MacLeod J.N., Minor R.R.;			
RT	"Sequence of normal canine COL1A1 cDNA and identification of a			
RT	heterozygous alpha 1(I) collagen Gly208Ala mutation in a severe case			
RT	of canine osteogenesis imperfecta.";			
RL	Arch. Biochem. Biophys. 384:37-46(2000).			
CC	-I- FUNCTION: Type I collagen is a member of group I collagen			
CC	(fibrillar forming collagen).			
CC	-I- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.			
CC	-I- PTM: Prolines at the third position of the tripeptide repeating			
CC	unit (G-X-Y) are hydroxylated in some or all of the chains.			
CC	-I- DISEASE: Defects in COL1A1 are a cause of osteogenesis imperfecta			
CC	(OI).			
CC	-I- SIMILARITY: Contains 1 VWFC domain.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; AF153362; AAD34619.1; -			
DR	InterPro; IPR008161; C1g_helix.			
DR	InterPro; IPR008160; Collagen.			
DR	InterPro; IPR000885; Fib.collagen_C.			
DR	InterPro; IPR002181; Fibrinogen_C.			
DR	InterPro; IPR001007; VWFC_C.			
DR	Pfam; PF01410; COLFI; 1.			
DR	Pfam; PF01391; Collagen; 18.			
DR	ProDom; PD000007; C1g_helix; 2.			
DR	ProDom; PD002078; Fib.collagen_C; 1.			
DR	SMART; SM00038; COLFI; 1.			
DR	SMART; SM00214; VWFC; 1.			
DR	PROSITE; PS01208; VWFC_1; 1.			
DR	PROSITE; PS01184; VWFC_2; 1.			
KW	Extracellular matrix; Connective tissue; Repeat; Hydroxylation;			
KW	Glycoprotein; Collagen; Signal; Disease mutation.			
FT	SIGNAL 1 22			
	BY SIMILARITY.			

34	571.5	48.3	684	1	CA39_HUMAN	Q14050 homo sapien
35	571	48.2	1745	1	CA35_HUMAN	P25940 homo sapien
36	568.5	48.0	1804	1	CA1B_MOUSE	Q61245 mus musculu
37	563.5	47.6	2944	1	CA17_HUMAN	Q02388 homo sapien
38	563	47.6	1685	1	CA54_HUMAN	P29400 homo sapien
39	561	47.4	1603	1	CA1F_HUMAN	Q07092 homo sapien
40	555.5	46.9	921	1	CA19_HUMAN	P20849 homo sapien
41	552.5	46.7	674	1	CA1A_CHICK	P08125 gallus gall
42	552.5	46.7	680	1	CA1A_HUMAN	Q03692 homo sapien
43	549	46.4	526	1	CA21_RABIT	Q28668 oryctolagus
44	548.5	46.3	921	1	CA19_MOUSE	Q05722 mus musculu
45	547	46.2	1712	1	CA24_HUMAN	P08572 homo sapien

FT PROPEP 23 157 AMINO-TERMINAL PROPEPTIDE.
 FT CHAIN 158 1214 COLLAGEN ALPHA 1(I) CHAIN.
 FT PROPEP 1215 1460 CARBOXYL-TERMINAL PROPEPTIDE.
 FT DOMAIN 34 92 VWFC.
 FT DOMAIN 158 174 NONHELICAL REGION (N-TERMINAL).
 FT DOMAIN 175 1188 TRIPLE-HELICAL REGION.
 FT SITE 1189 1214 NONHELICAL REGION (C-TERMINAL).
 FT SITE 741 743 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 1089 1091 CELL ATTACHMENT SITE (POTENTIAL).
 FT CARBOHYD 1361 1361 N-LINKED (GLCNAC...) (POTENTIAL).
 FT VARIANT 208 208 G -> A (in OI, severe).
 SQ SEQUENCE 1460 AA; 138762 MW; 58E3674D2B570697 CRC64;

Query Match 97.0%; Score 1149; DB 1; Length 1460;
 Best Local Similarity 93.6%; Pred. No. 4.1e-48;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPGGPGGSGRGGPGGAGVAGKPGAGRGSGPPA 48
 |||||
 Db 454 GPPGAGGEGKRGARGEPTGLPGPPGGERGGSGRGGPGGAGVAGKPGAGRGSGPPA 513
 |||||

QY 49 GPKSGPGEAGRGEAGLFGKAGLTGSGPGPGDGKTPPGAGQDGRPPGPPGARGQA 108
 |||||

Db 514 GPKSGPGEAGRGEAGLFGKAGLTGSGPGPGDGKTPPGAGQDGRPPGPPGARGQA 573
 |||||

QY 109 GVMGPPGKGAAGPCKAGRGVPGPGAVCPAGKDGAGAGQPPGAGPAGRGEGGPA 168
 |||||

Db 574 GVMGPPGKGAAGPCKAGRGVPGPGAVCPAGKDGAGAGQPPGAGPAGRGEGGPA 633
 |||||

QY 169 GSPFGQLPGPAGPGGAGKGEQGVCDLCAAPGSPGAG 208
 |||||

Db 634 GSPFGQLPGPAGPGEAGKGEQGVCDLCAAPGSPGAG 673
 |||||

RESULT 2
 CALL HUMAN
 ID CALL HUMAN STANDARD; PRT; 1464 AA.
 AC P02452; Q14037; Q15176;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Collagen alpha 1(I) chain precursor.
 GN COL1A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE OF 1-472 FROM N.A.
 RX MEDLINE=89025644; PubMed=3178743;
 RA Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
 RA Jaenisch R., Prockop D.J.;
 RT "Structure of a full-length cDNA clone for the prepro alpha 1(I)
 RT chain of human type I procollagen."
 RL Biochem. J. 253:919-922(1988).
 RN [2]
 RP SEQUENCE OF 1-181 FROM N.A.
 RX MEDLINE=84270697; PubMed=6462220;
 RA Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morabito M.,
 RA Myers J., Williams C., Ramirez F.;
 RT "Human pro alpha 1(I) collagen gene structure reveals evolutionary
 RT conservation of a pattern of introns and exons."
 RL Nature 310:337-340(1984).
 RN [3]
 RP SEQUENCE OF 162-301.
 RC TISSUE=Skin;
 RX MEDLINE=71038625; PubMed=5529814;
 RA Click E.M., Bornstein P.;
 RT "Isolation and characterization of the cyanogen bromide peptides from
 RT the alpha 1 and alpha 2 chains of human skin collagen."
 RL Biochemistry 9:4699-4706(1970).
 RN [4]
 RP SEQUENCE OF 263-268.

RC TISSUE=Skin;
 RX MEDLINE=71001508; PubMed=4319110;
 RA Morgan P.H., Jacobs H.G., Segrest J.P., Cunningham L.W.;
 RT "A comparative study of glycopeptides derived from selected
 RT vertebrate collagens. A possible role of the carbohydrate in fibril
 RT formation."
 RL J. Biol. Chem. 245:5042-5048(1970).
 RN [5]
 RP SEQUENCE OF 425-1464 FROM N.A.
 RX MEDLINE=84080385; PubMed=6689127;
 RA Bernard M.P., Chu M.-L., Myers J.C., Ramirez F., Eikenberry E.F.,
 RA Prockop D.J.;
 RT "Nucleotide sequences of complementary deoxyribonucleic acids for the
 RT pro alpha 1 chain of human type I procollagen. Statistical evaluation
 RT of structures that are conserved during evolution."
 RL Biochemistry 22:5213-5223(1983).
 RN [6]
 RP SEQUENCE OF 1229-1454 FROM N.A.
 RC TISSUE=Bone;
 RX MEDLINE=88124208; PubMed=3340531;
 RA Maekelae J.K., Raessina M., Virta A., Vuorio E.;
 RT "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide
 RT domain."
 RL Nucleic Acids Res. 16:349-349(1988).
 RN [7]
 RP SEQUENCE OF 1-34 FROM N.A.
 RX MEDLINE=88097389; PubMed=3480516;
 RA Bornstein P., McKay J., Morishima J.K., Devarayalu S., Gelinas R.E.;
 RT "Regulatory elements in the first intron contribute to
 RT transcriptional control of the human alpha 1(I) collagen gene."
 RL Proc. Natl. Acad. Sci. U.S.A. 84:8869-8873(1987).
 RN [8]
 RP SEQUENCE OF 1-34 FROM N.A.
 RX MEDLINE=85130970; PubMed=2857713;
 RA Chu M.-L., de Wet W.J., Bernard M.P., Ramirez F.;
 RT "Fine structural analysis of the human pro-alpha 1(I) collagen gene.
 RT Promoter structure, AluI repeats, and polymorphic transcripts."
 RL J. Biol. Chem. 260:2315-2320(1985).
 RN [9]
 RP SEQUENCE OF 1-44 FROM N.A.
 RX MEDLINE=88033098; PubMed=2822714;
 RA Rossouw C.M.S., Vergeer W.P., du Plooy S.J., Bernard M.P., Ramirez F.,
 RA de Wet W.J.;
 RT "DNA sequences in the first intron of the human pro-alpha 1(I)
 RT collagen gene enhance transcription."
 RL J. Biol. Chem. 262:15151-15157(1987).
 RN [10]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=91184577; PubMed=2010058;
 RA Kuivaniemi H., Tromp G., Prockop D.J.;
 RT "Mutations in collagen genes: causes of rare and some common diseases
 RT in humans."
 RL FASEB J. 5:2052-2060(1991).
 RN [11]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=97255959; PubMed=9101290;
 RA Kuivaniemi H., Tromp G., Prockop D.J.;
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
 RT associated collagen (type IX), and network-forming collagen (type X)
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels."
 RL Hum. Mutat. 9:300-315(1997).
 RN [12]
 RP REVIEW ON OI VARIANTS.
 RX MEDLINE=91374476; PubMed=1895312;
 RA Byers P.H., Wallis G.A., Willing M.C.;
 RT "Osteogenesis imperfecta: translation of mutation to phenotype."
 RL J. Med. Genet. 28:433-442(1991).
 RN [13]
 RP REVIEW ON OI VARIANTS.
 RX MEDLINE=97169389; PubMed=9016532;
 RA Daigleish R.;
 RT "The human type I collagen mutation database."
 RL Nucleic Acids Res. 25:181-187(1997).

RL J. Biol. Chem. 242:2572-2574 (1967).
 RN [4]
 RP SEQUENCE OF 56-102.
 RX MEDLINE=71263178; PubMed=43273399;
 RA Butler W.T., Ponds S.L.;
 RT "Chemical studies on the cyanogen bromide peptides of rat skin
 collagen. Amino acid sequence of alpha 1-CB4.";
 RL Biochemistry 10:2076-2081 (1971).
 RN [5]
 RP SEQUENCE OF 103-139.
 RX MEDLINE=70085124; PubMed=5411206;
 RA Butler W.T.;
 RT "Chemical studies on the cyanogen bromide peptides of rat skin
 collagen. The covalent structure of alpha 1-CB5, the major
 hexose-containing cyanogen bromide peptide of alpha 1.";
 RL Biochemistry 9:44-50 (1970).
 RN [6]
 RP SEQUENCE OF 140-238.
 RX MEDLINE=72136131; PubMed=4335087;
 RA Balian G., Click E.M., Bornstein P.;
 RT "Structure of rat skin collagen alpha 1-CB8. Amino acid sequence of
 the hydroxylamine-produced fragment HA1.";
 RL Biochemistry 10:4470-4478 (1971).
 RN [7]
 RP SEQUENCE OF 239-418.
 RX MEDLINE=73006942; PubMed=4342027;
 RA Balian G., Click E.M., Hermodson M.A., Bornstein P.;
 RT "Structure of rat skin collagen alpha 1-CB8. Amino acid sequence of
 the hydroxylamine-produced fragment HA2.";
 RL Biochemistry 11:3798-3806 (1972).
 RN [8]
 RP SEQUENCE OF 419-567.
 RX MEDLINE=74271984; PubMed=4366532;
 RA Bulter W.T., Underwood S.P., Finch J.E. Jr.;
 RT "Chemical studies on the cyanogen bromide peptides of rat skin
 collagen. Amino acid sequence of alpha 1-CB3.";
 RL Biochemistry 13:2946-2953 (1974).
 RN [9]
 RP SEQUENCE OF 568-651.
 RX MEDLINE=74011954; PubMed=4126850;
 RA Stoltz M., Timpi R., Furtmayr H., Kuehn K.;
 RT "Structural and immunogenic properties of a major antigenic
 determinant in neutral salt-extracted rat-skin collagen.";
 RL Eur. J. Biochem. 37:287-294 (1973).
 RN [10]
 RP SEQUENCE OF 651-671.
 RX MEDLINE=73049495; PubMed=4636751;
 RA Stoltz M., Timpi R., Kuehn K.;
 RT "Non-helical regions in rat collagen alpha 1-chain.";
 RL FEBS Lett. 26:61-65 (1972).
 RN [11]
 RP SEQUENCE OF 529-567 FROM N.A.
 RX MEDLINE=85122694; PubMed=6395893;
 RA Genovese C., Rowe D., Kream B.;
 RT "Construction of DNA sequences complementary to rat alpha 1 and alpha
 2 collagen mRNA and their use in studying the regulation of type I
 collagen synthesis by 1,25-dihydroxyvitamin D.";
 RL Biochemistry 23:6210-6216 (1984).
 CC -!- FUNCTION: Type I collagen is a member of group I collagen
 (fibrillar forming collagen).
 CC -!- SUBUNIT: Trimer of one alpha 2(I) and two alpha 1(I) chains.
 CC -!- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
 bones. In bones the fibrils are mineralized with calcium
 hydroxyapatite.
 CC -!- PTM: Proline residues at the third position of the tripeptide
 repeating unit (G-X-Y) are hydroxylated in some or all of the
 chains.
 CC -!- PTM: O-linked glycan consists of a Glc-Gal disaccharide bound to
 the oxygen atom of a post-translationally added hydroxyl group.
 CC -----
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M1432; AAA40832.1; ALT SEQ.
 DR InterPro: IPR008161; Clg_helix.
 DR InterPro: IPR008160; Collagen.
 DR InterPro: IPR001007; VWF_C.
 DR ProDom: PD000007; Clg_helix; 1.
 DR PROSITE: PS01208; VWF_C1; PARTIAL.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen; Pyrrolidone carboxylic acid.
 FT MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID (PROBABLE).
 FT MOD_RES 9 9 CONVERTED TO AN ALDEHYDE GROUP THAT IS
 FT INVOLVED IN CROSS-LINKING.
 FT MOD_RES 28 28 HYDROXYLATION (PROBABLE).
 FT MOD_RES 31 31 HYDROXYLATION (PROBABLE).
 FT MOD_RES 34 34 HYDROXYLATION (PROBABLE).
 FT MOD_RES 43 43 HYDROXYLATION (PROBABLE).
 FT MOD_RES 46 46 HYDROXYLATION (PROBABLE).
 FT MOD_RES 49 49 HYDROXYLATION (PROBABLE).
 FT MOD_RES 103 103 HYDROXYLATION (PROBABLE).
 FT CARBOHYD 103 103 O-LINKED (GAL.).
 FT MOD_RES 124 124 HYDROXYLATION (PROBABLE).
 FT MOD_RES 424 424 HYDROXYLATION (PROBABLE).
 FT MOD_RES 547 547 HYDROXYLATION (PROBABLE).
 FT NON_CONS 567 568
 FT DOMAIN 641 651
 FT SEQUENCE 671 AA; 50615 NW; 9DC3114204AC4918 CRC64;
 SQ
 Query Match 94.4%; Score 1118; DB 1; Length 671;
 Best Local Similarity 90.0%; Pred. No. 7e-47;
 Matches 198; Conservative 5; Mismatches 5; Indels 12; Gaps 1;
 QY 1 GPP-----GEPGTGLPGPPGCGSGRPGADGVAGPKGPPGERSGCPA 48
 DB 296 GPPGAGREGKRGARCEGPGSLGPGGCGSGRPGADGVAGPKGPPGERSGCPA 355
 QY 49 GPKSGPAGRPGEAGLPGAKLGTSPGSGPDGKTGPPGAGQDGRPPGPPGARGQA 108
 DB 356 GPKSGPAGRPGEAGLPGAKLGTSPGSGPDGKTGPPGAGZGCRPPGPPGARGQA 415
 QY 109 GVMGPPGPKGAAGSGKAGRGVPGPCAVGPKGDEAGGAGGQGGGPPGPPGARGGEOGA 168
 DB 416 GVMGPPGPKGATAGTGGKAGRGVPGPCAVGPKGDEAGGAGGQGGGPPGPPGARGGEOGA 475
 QY 169 GSPGFGGLPGPAGPFGGAGKFGEGQGVFGDLGAPGSGPAG 208
 DB 476 GSPGFGGLPGPAGPFGGAGKFGGZGVFGDLGAPGSGARG 515
 RESULT 4
 CALL MOUSE
 ID CALL MOUSE STANDARD; PRT; 1453 AA.
 AC P11087; Q60635;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Collagen alpha 1(I) chain precursor.
 GN COL1A1 OR COL1A.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]_TaxID=10090;
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N;
 RX MEDLINE=96033240; PubMed=8535610;
 RA Li S.W., Khilian J., Prockop D.J.;
 RT "The complete cDNA coding sequence for the mouse pro alpha 1(I) chain
 of type I procollagen.";
 RL Matrix Biol. 14:593-595 (1995).

[2]
 R1 PROSITE, P850184; VMFC_2; 1.
 R2 Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 R3 Glycoprotein; Collagen; Signal.
 R4 SIGNAL 1 22
 R5 FT PROPEP 23 151 AMINO-TERMINAL PROPEPTIDE.
 R6 FT CHAIN 152 1207 COLLAGEN ALPHA 1(I) CHAIN.
 R7 FT PROPEP 1208 1453 CARBOXYL-TERMINAL PROPEPTIDE.
 R8 FT DOMAIN 29 87 VMFC.
 R9 FT DOMAIN 152 167 NONHELICAL REGION (N-TERMINAL).
 R10 FT DOMAIN 158 1181 TRIPLE-HELICAL REGION.
 R11 FT DOMAIN 1182 1207 NONHELICAL REGION (C-TERMINAL).
 R12 FT CARBOHYD 56 56 N-LINKED (GLCNAC. .) (POTENTIAL).
 R13 FT CARBOHYD 1354 1354 N-LINKED (GLCNAC. .) (POTENTIAL).
 R14 FT SITE 734 736 CELL ATTACHMENT SITE (POTENTIAL).
 R15 FT SITE 1082 1084 CELL ATTACHMENT SITE (POTENTIAL).
 R16 FT CONFLICT 1450 1450 A -> V (IN REF. 5).
 R17 SQ SEQUENCE 1453 AA; 137944 MW; 3802535DF81808 CRC64;
 R18
 R19 Query Match 94.3%; Score 1116; DB 1; Length 1453;
 R20 Best Local Similarity 90.5%; Pred. No. 1.5e-46;
 R21 Matches 199; Conservative 4; Mismatches 5; Indels 12; Gaps 1;
 R22
 R23 QY 1 GPP-----GEPGPTGLPGPGRGPGSGRGGPGGADGVAGPKGAGRGSPGPA 48
 R24 DB 447 GPPGAGEGKRGARGPGLPGPPGGRGGPGGADGVAGPKGSGERGAPGPA 506
 R25 QY 49 GPKGSGEAGRPCEAGLPCAKGLTSGPSGPDGKTGTPPGAGQDGRPGPPGARGQA 108
 R26 DB 507 GPKGSGEAGRPCEAGLPCAKGLTSGPSGPDGKTGTPPGAGQDGRPGPPGARGQA 566
 R27 QY 109 GVGFPGPAGAGPKAGRGVPPGAVGAGKDGAGAGQDGRPGAGRGEGQGPA 168
 R28 DB 567 GVGFPGPAGAGPKAGRGVPPGAVGAGKDGAGAGQDGRPGAGRGEGQGPA 626
 R29 QY 169 GSPGFGPLPGPAGPCEAGKPGCEGQGVPGDLGAPGSGGAPG 208
 R30 DB 627 GSPGFGPLPGPAGPCEAGKPGCEGQGVPGDLGAPGSGGAPG 666
 R31
 R32 RESULT 5
 R33 CALL_CHICK STANDARD; PRT; 1453 AA.
 R34 AC P02457; 1986 (Rel. 01, Created)
 R35 DT 21-JUL-1989 (Rel. 12, Last sequence update)
 R36 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 R37 DB Collagen alpha 1(I) chain precursor.
 R38 GN COL1A1.
 R39 OS Gallus gallus (Chicken).
 R40 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 R41 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 R42 OC Gallus.
 R43 CX NCBI_TaxID=9031;
 R44 RN [1]
 R45 RP SEQUENCE OF 1-153 FROM N.A.
 R46 RX MEDLINE=88056316; PubMed=3678834;
 R47 RA Finer M.H., Boedtker H., Doty P.;
 R48 RT "Construction and characterization of cDNA clones encoding the 5' end
 R49 of the chicken pro alpha 1(I) collagen mRNA.";
 R50 RL J. Biol. Chem. 262:13323-13332(1987).
 R51 RN [2]
 R52 RP SEQUENCE OF 1-144 FROM N.A.
 R53 RX MEDLINE=88007542; PubMed=2820966;
 R54 RA Finer M.H., Aho S., Garstenfeld L.C., Boedtker H., Doty P.;
 R55 RT "Unusual DNA sequences located within the promoter region and the
 R56 first intron of the chicken pro-alpha 1(I) collagen gene.";
 R57 RL J. Biol. Chem. 262:13323-13332(1987).
 R58 RN [3]
 R59 RP SEQUENCE OF 152-1187.
 R60 RX MEDLINE=82231995; PubMed=7093229;
 R61 RA Higberger J.H., Corbett C., Dixit S.N., Yu W., Seyer J.M.,
 R62 Kang A.H., Gross J.;
 R63 RT "Amino acid sequence of chick skin collagen alpha 1(I)-CB8 and the

R100 MEDLINE=86137403; PubMed=3841523;
 R101 French B.T., Lee W.-H., Maul G.G.;
 R102 "Nucleotide sequence of a cDNA clone for mouse pro alpha 1(I)
 R103 collagen protein.";
 R104 Gene 39:311-312(1985).
 R105 [3]
 R106 R107 R108 R109 R110 R111 R112 R113 R114 R115 R116 R117 R118 R119 R120 R121 R122 R123 R124 R125 R126 R127 R128 R129 R130 R131 R132 R133 R134 R135 R136 R137 R138 R139 R140 R141 R142 R143 R144 R145 R146 R147 R148 R149 R150 R151 R152 R153 R154 R155 R156 R157 R158 R159 R160 R161 R162 R163 R164 R165 R166 R167 R168 R169 R170 R171 R172 R173 R174 R175 R176 R177 R178 R179 R180 R181 R182 R183 R184 R185 R186 R187 R188 R189 R190 R191 R192 R193 R194 R195 R196 R197 R198 R199 R200 R201 R202 R203 R204 R205 R206 R207 R208 R209 R210 R211 R212 R213 R214 R215 R216 R217 R218 R219 R220 R221 R222 R223 R224 R225 R226 R227 R228 R229 R230 R231 R232 R233 R234 R235 R236 R237 R238 R239 R240 R241 R242 R243 R244 R245 R246 R247 R248 R249 R250 R251 R252 R253 R254 R255 R256 R257 R258 R259 R260 R261 R262 R263 R264 R265 R266 R267 R268 R269 R270 R271 R272 R273 R274 R275 R276 R277 R278 R279 R280 R281 R282 R283 R284 R285 R286 R287 R288 R289 R290 R291 R292 R293 R294 R295 R296 R297 R298 R299 R300 R301 R302 R303 R304 R305 R306 R307 R308 R309 R310 R311 R312 R313 R314 R315 R316 R317 R318 R319 R320 R321 R322 R323 R324 R325 R326 R327 R328 R329 R330 R331 R332 R333 R334 R335 R336 R337 R338 R339 R340 R341 R342 R343 R344 R345 R346 R347 R348 R349 R350 R351 R352 R353 R354 R355 R356 R357 R358 R359 R360 R361 R362 R363 R364 R365 R366 R367 R368 R369 R370 R371 R372 R373 R374 R375 R376 R377 R378 R379 R380 R381 R382 R383 R384 R385 R386 R387 R388 R389 R390 R391 R392 R393 R394 R395 R396 R397 R398 R399 R400 R401 R402 R403 R404 R405 R406 R407 R408 R409 R410 R411 R412 R413 R414 R415 R416 R417 R418 R419 R420 R421 R422 R423 R424 R425 R426 R427 R428 R429 R430 R431 R432 R433 R434 R435 R436 R437 R438 R439 R440 R441 R442 R443 R444 R445 R446 R447 R448 R449 R450 R451 R452 R453 R454 R455 R456 R457 R458 R459 R460 R461 R462 R463 R464 R465 R466 R467 R468 R469 R470 R471 R472 R473 R474 R475 R476 R477 R478 R479 R480 R481 R482 R483 R484 R485 R486 R487 R488 R489 R490 R491 R492 R493 R494 R495 R496 R497 R498 R499 R500 R501 R502 R503 R504 R505 R506 R507 R508 R509 R510 R511 R512 R513 R514 R515 R516 R517 R518 R519 R520 R521 R522 R523 R524 R525 R526 R527 R528 R529 R530 R531 R532 R533 R534 R535 R536 R537 R538 R539 R540 R541 R542 R543 R544 R545 R546 R547 R548 R549 R550 R551 R552 R553 R554 R555 R556 R557 R558 R559 R560 R561 R562 R563 R564 R565 R566 R567 R568 R569 R570 R571 R572 R573 R574 R575 R576 R577 R578 R579 R580 R581 R582 R583 R584 R585 R586 R587 R588 R589 R590 R591 R592 R593 R594 R595 R596 R597 R598 R599 R600 R601 R602 R603 R604 R605 R606 R607 R608 R609 R610 R611 R612 R613 R614 R615 R616 R617 R618 R619 R620 R621 R622 R623 R624 R625 R626 R627 R628 R629 R630 R631 R632 R633 R634 R635 R636 R637 R638 R639 R640 R641 R642 R643 R644 R645 R646 R647 R648 R649 R650 R651 R652 R653 R654 R655 R656 R657 R658 R659 R660 R661 R662 R663 R664 R665 R666 R667 R668 R669 R670 R671 R672 R673 R674 R675 R676 R677 R678 R679 R680 R681 R682 R683 R684 R685 R686 R687 R688 R689 R690 R691 R692 R693 R694 R695 R696 R697 R698 R699 R700 R701 R702 R703 R704 R705 R706 R707 R708 R709 R710 R711 R712 R713 R714 R715 R716 R717 R718 R719 R720 R721 R722 R723 R724 R725 R726 R727 R728 R729 R730 R731 R732 R733 R734 R735 R736 R737 R738 R739 R740 R741 R742 R743 R744 R745 R746 R747 R748 R749 R750 R751 R752 R753 R754 R755 R756 R757 R758 R759 R760 R761 R762 R763 R764 R765 R766 R767 R768 R769 R770 R771 R772 R773 R774 R775 R776 R777 R778 R779 R780 R781 R782 R783 R784 R785 R786 R787 R788 R789 R790 R791 R792 R793 R794 R795 R796 R797 R798 R799 R800 R801 R802 R803 R804 R805 R806 R807 R808 R809 R810 R811 R812 R813 R814 R815 R816 R817 R818 R819 R820 R821 R822 R823 R824 R825 R826 R827 R828 R829 R830 R831 R832 R833 R834 R835 R836 R837 R838 R839 R840 R841 R842 R843 R844 R845 R846 R847 R848 R849 R850 R851 R852 R853 R854 R855 R856 R857 R858 R859 R860 R861 R862 R863 R864 R865 R866 R867 R868 R869 R870 R871 R872 R873 R874 R875 R876 R877 R878 R879 R880 R881 R882 R883 R884 R885 R886 R887 R888 R889 R890 R891 R892 R893 R894 R895 R896 R897 R898 R899 R900 R901 R902 R903 R904 R905 R906 R907 R908 R909 R910 R911 R912 R913 R914 R915 R916 R917 R918 R919 R920 R921 R922 R923 R924 R925 R926 R927 R928 R929 R930 R931 R932 R933 R934 R935 R936 R937 R938 R939 R940 R941 R942 R943 R944 R945 R946 R947 R948 R949 R950 R951 R952 R953 R954 R955 R956 R957 R958 R959 R960 R961 R962 R963 R964 R965 R966 R967 R968 R969 R970 R971 R972 R973 R974 R975 R976 R977 R978 R979 R980 R981 R982 R983 R984 R985 R986 R987 R988 R989 R990 R991 R992 R993 R994 R995 R996 R997 R998 R999

RX MEDLINE=89231683; PubMed=2714276;
 RA Sever J.M., Hasty K.A., Kang A.H.;
 RT "Covalent structure of collagen. Amino acid sequence of an
 RT arthrogenic cyanogen bromide peptide from type II collagen of
 RL bovine cartilage.";
 RL Eur. J. Biochem. 181:159-173 (1989).
 [5]
 RP SEQUENCE OF 418-492.
 RX MEDLINE=74163168; PubMed=4857180;
 RA Butler W.T., Miller E.J., Finch J.E. Jr., Inagami T.;
 RT "Homologous regions of collagen alpha(I) and alpha(II) chains:
 RT apparent clustering of variable and invariant amino acid residues.";
 RL Biochem. Biophys. Res. Commun. 57:190-195 (1974).
 [6]
 RP SEQUENCE OF 180-272 FROM N.A.
 RC TISSUE=Cartilage;
 RX MEDLINE=94194070; PubMed=7511638;
 RA Brand D.D., Myers L.K., Terato K., Whittington K.B., Stuart J.M.,
 RA Rosloniec E.F.;
 RT "Characterization of the T cell determinants in the induction of
 RT autoimmune arthritis by bovine alpha 1(II)-CB11 in H-2q mice.";
 RL J. Immunol. 152:3088-3097 (1994).
 [7]
 RP SEQUENCE OF 417-566 FROM N.A.
 RC TISSUE=Cartilage;
 RX MEDLINE=99410731; PubMed=10479530;
 RA Tang B., Chiang T.M., Brand D.D., Gumanovskaya M.L., Stuart J.M.,
 RA Kang A.H., Myers L.K.;
 RT "Molecular definition and characterization of recombinant bovine CB9
 RT and CB10: immunogenicity and arthritogenicity.";
 RL Clin. Immunol. 92:256-264 (1999).
 [8]
 RP SEQUENCE OF 567-747 FROM N.A.
 RX MEDLINE=85215651; PubMed=2582365;
 RA Sangiorgi F.O., Benson-Chanda V., de Wet W.J., Sobel M.E.,
 RA Ramirez F.;
 RT "Analysis of cDNA and genomic clones coding for the pro alpha 1 chain
 RT of calf type II collagen.";
 RL Nucleic Acids Res. 13:2815-2826 (1985).
 CC -!- FUNCTION: Collagen type II is specific for cartilaginous tissues.
 CC -!- SUBUNIT: Homotrimers of alpha 1(II) chains.
 CC -!- PTM: Proline residues at the third position in some or all of the
 CC repeating unit (G-X-Y) are hydroxylated in some or all of the
 CC chains. Hydroxylation on Pro-9 is involved in cross-linking.
 CC -!- PTM: O-linked glycans consist of Glc-Gal disaccharides bound to
 CC the oxygen atom of post-translationally added hydroxyl groups.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL: L28918; AAA30436.1; --
 DR EMBL: AF138957; A242347.1; --
 DR EMBL: X02420; CAB26269.1; --
 DR PIR: A90363; CAB08C.
 DR PIR: I45876; I45876.
 DR InterPro: IPR008161; Clg_helix.
 DR InterPro: IPR008160; Collagen.
 DR InterPro: IPR000885; Fib_collagen_C.
 DR InterPro: IPR001007; VFV_C.
 DR ProDom: PD000007; Clg_helix; 2.
 DR ProDom: PD002078; Fib_collagen_C; 1.
 DR SMART: SM00038; COLFI; 1.
 DR PROSITE: PS01208; VFV_C; PARTIAL.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen.
 FT CHAIN 1 >566 COLLAGEN ALPHA 1(II) CHAIN.
 FT NON CONS 566 567
 FT PROPEP <567 747 CARBOXYL-TERMINAL PROPEPTIDE.

FT MOD RES 9 9
 FT MOD RES 102 102
 FT CARBOHYD 102 102
 FT MOD RES 114 114
 FT CARBOHYD 114 114
 FT MOD RES 123 123
 FT CARBOHYD 123 123
 FT MOD RES 189 189
 FT CARBOHYD 189 189
 FT MOD RES 423 423
 FT CARBOHYD 423 423
 FT MOD RES 435 435
 FT CARBOHYD 435 435
 FT VARIANT 143 143
 FT VARIANT 164 164
 FT CONFLICT 179 179
 FT CONFLICT 185 185
 FT CONFLICT 191 191
 FT CONFLICT 195 195
 FT CONFLICT 215 215
 FT CONFLICT 227 227
 FT CONFLICT 251 251
 FT CONFLICT 258 258
 FT CONFLICT 261 261
 FT CONFLICT 492 492
 SQ SEQUENCE 747 AA; 71329 MW; DQFCD7CDICAF77C CRC64;
 Query Match 77.1%; Score 913; DB 1; Length 747;
 Best Local Similarity 78.4%; Pred. No. 3.6e-37;
 Matches 163; Conservative 6; Mismatches 39; Indels 0; Gaps 0;
 QY 1 GPPGEPGPTGLPGPPGCGSGPSRGGAGDGVAGPKGAPAGERGSGPGAGPKSGPGAGRP 60
 DB 307 GARGEPGGAGPAGPPGCGGAGPSRGGAGDGVAGPKGAPAGERGSGPGAGPKSGPGAGRP 366
 QY 61 GEAGLFGAAXGLTSGPSGPDGKHTGTPGAGQDGRPPGPPGPPGPPGPPGPPGPPGPPG 120
 DB 367 GEAGLFGAAXGLTGRPGDAGPQGVGSGAGPFGDGRPPGPPGPPGPPGPPGPPGPPG 426
 QY 121 GEPKAGGVRGPPGAVGAGPKGAGGAGGAGGPPGPPGPPGPPGPPGPPGPPGPPGPPG 180
 DB 427 GEPKAGGKGLPGAGLGLPGDGTGAAGPPGPPGPPGPPGPPGPPGPPGPPGPPG 486
 QY 181 GPPGAGKPGGQGVGDLGAPGSPGAG 208
 DB 487 GPPGEGKPGDQGVGEGAGAPGLVGRG 514
 RESULT 7
 CAL12 HUMAN STANDARD; PRT; 1418 AA.
 ID CAL12 HUMAN
 AC P02458;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Collagen alpha 1(II) chain precursor [Contains: Chondrocalcin].
 GN COL2A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBT_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90067946; PubMed=2587267;
 RA Su M.W., Lee B., Ramirez F., Machado M., Horton W.;
 RT "Nucleotide sequence of the full length cDNA encoding for human type
 RT II procollagen.";
 RL Nucleic Acids Res. 17:9473-9473 (1989).
 [2]
 RP SEQUENCE OF 1-28 FROM N.A.
 RX MEDLINE=87031574; PubMed=3021582;
 RA Nunez A.M., Kohno K., Martin G.R., Yamada Y.;
 RT "Promoter region of the human pro-alpha 1(II)-collagen gene.";
 RL Gene 44:11-16 (1986).

RN [3] SEQUENCE OF 432-1145 FROM N.A.
RA Ramirez F.;
RL Submitted (DEC-1988) to the EMBL/GenBank/DBJ databases.
RN [4]
RX SEQUENCE OF 963-1418 FROM N.A.
RA MEDLINE=85190334; PubMed=3857598;
RA Chan K.S.E.; Stoker N.G.; Griffin J.R.; Grosfeld F.G.; Solomon E.;
RT "Identification and characterization of the human type II collagen
RT gene (COL2A1).";
RL Proc. Natl. Acad. Sci. U.S.A. 82:2555-2559(1985).
RN [5]
RX SEQUENCE OF 1120-1398 FROM N.A.
RA MEDLINE=85306861; PubMed=3840017;
RA Elina K.; Mäkelä J.K.; Vuorio T.; Kauppinen S.; Knowles J.;
RA Vuorio E.;
RT "Construction and identification of a cDNA clone for human type II
RT procollagen mRNA.";
RL Biochem. J. 229:183-188(1985).
RN [6]
RX SEQUENCE OF 1106-1418 FROM N.A.
RA MEDLINE=89067771; PubMed=2825137;
RA Elina K.; Vuorio T.; Vuorio E.;
RT "Determination of the single polyadenylation site of the human pro
RT alpha 1(II) collagen gene.";
RL Nucleic Acids Res. 15:9499-9504(1987).
RN [7]
RX SEQUENCE OF 1227-1289 FROM N.A.
RA MEDLINE=86104139; PubMed=3002437;
RA Nunez A.M.; Francomano C.; Young M.F.; Martin G.R.; Yamada Y.;
RT "Isolation and partial characterization of genomic clones coding for
RT a human pro-alpha 1 (II) collagen chain and demonstration of
RT restriction fragment length polymorphism at the 3' end of the gene.";
RL Biochemistry 24:6343-6348(1985).
RN [8]
RX SEQUENCE OF 1176-1226 FROM N.A.
RA MEDLINE=84118798; PubMed=6320112;
RA Strom C.M.; Upholt W.B.;
RT "Isolation and characterization of genomic clones corresponding to
RT the human type II procollagen gene.";
RL Nucleic Acids Res. 12:1025-1038(1984).
RN [9]
RX SEQUENCE OF 35-167 FROM N.A.
RA MEDLINE=89233138; PubMed=2714801;
RA Su M.W.; Benson-Chanda V.; Vissing H.; Ramirez F.;
RT "Organization of the exons coding for pro alpha 1(II) collagen N-
RT propeptide confirms a distinct evolutionary history of this domain of
RT the fibrillar collagen genes.";
RL Genomics 4:438-441(1989).
RN [10]
RX REVIEW ON VARIANTS.
RA MEDLINE=91184577; PubMed=2010058;
RA Kuivaniemi H.; Tromp G.; Prockop D.J.;
RT "Mutations in collagen genes: causes of rare and some common diseases
RT in humans.";
RL PASEB J. 5:2052-2060(1991).
RN [11]
RX REVIEW ON VARIANTS.
RA MEDLINE=97255959; PubMed=9101290;
RA Kuivaniemi H.; Tromp G.; Prockop D.J.;
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT associated collagen (type IX), and network-forming collagen (type X)
RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL Hum. Mutat. 9:300-315(1997).
RN [12]
RX VARIANT SER-1074.
RA MEDLINE=90036909; PubMed=2572591;
RA Vissing H.; D'Alessio M.; Lee B.; Ramirez F.; Godfrey M.;
RA Hollister D.W.;
RT "Glycine to serine substitution in the triple helical domain of pro-
RT alpha 1(II) collagen results in a lethal perinatal form of short-
RT limbed dwarfism.";
RL J. Biol. Chem. 264:18265-18267(1989).
RN [13]
RX VARIANT SEDC 1095-GLY--TYR-1330 DEL.
RA MEDLINE=89266907; PubMed=2543071;
RA Lee B.; Vissing H.; Ramirez F.; Rogers D.; Rimoin D.;
RT "Identification of the molecular defect in a family with
RT spondyloepiphyseal dysplasia.";
RL Science 244:978-980(1989).
RN [14]
RX VARIANT OSTEOARTHRITIS CYS-650.
RA MEDLINE=90370826; PubMed=1975693;
RA Ala-Kokko L.; Baldwin C.T.; Moskowitz R.W.; Prockop D.J.;
RT "Single base mutation in the type II procollagen gene (COL2A1) as a
RT cause of primary osteoarthritis associated with a mild
RT chondrodysplasia.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:6565-6568(1990).
RN [15]
RX VARIANT OI-IV VAL-717.
RA MEDLINE=91291136; PubMed=2064612;
RA Bateman J.F.; Hannagan M.; Chan D.; Cole W.G.;
RT "Characterization of a type I collagen alpha 2(I) glycine-586 to
RT valine substitution in osteogenesis imperfecta type IV. Detection of
RT the mutation and prenatal diagnosis by a chemical cleavage method.";
RL Biochem. J. 276:765-770(1991).
RN [16]
RX VARIANT OSTEOARTHRITIS CYS-650.
RA MEDLINE=91086471; PubMed=1985108;
RA Eyre D.R.; Weis M.A.; Moskowitz R.W.;
RT "Cartilage expression of a type II collagen mutation in an inherited
RT form of osteoarthritis associated with a mild chondrodysplasia.";
RL J. Clin. Invest. 87:357-361(1991).
RN [17]
RX VARIANT HYPOCHONDROGENESIS GLU-984.
RA MEDLINE=93054548; PubMed=1429602;
RA Bogaert R.; Tiller G.E.; Wies M.A.; Gruber H.E.; Rimoin D.L.;
RA Cohn D.H.; Eyre D.R.;
RT "An amino acid substitution (Gly983-->Glu) in the collagen alpha
RT 1(II) chain produces hypochondrogenesis.";
RL J. Biol. Chem. 267:22523-22526(1992).
RN [18]
RX VARIANT HYPOCHONDROGENESIS SER-705.
RA MEDLINE=92262484; PubMed=1374906;
RA Horton W.A.; Machado M.A.; Ellard J.; Campbell D.; Bartley J.;
RA Ramirez F.; Vitale E.; Lee B.;
RT "Characterization of a type II collagen gene (COL2A1) mutation
RT identified in cultured chondrocytes from human hypochondrogenesis";
RL Proc. Natl. Acad. Sci. U.S.A. 89:4583-4587(1992).
RN [19]
RX VARIANT WS-II ASP-198.
RA MEDLINE=93304428; PubMed=8317498;
RA Koerkoe J.; Ritvaniemi P.; Haataja L.; Kaeaeiaieinen H.;
RA Kivirikko K.I.; Prockop D.J.; Ala-Kokko L.;
RT "Mutation in type II procollagen (COL2A1) that substitutes aspartate
RT for glycine alpha 1-67 and that causes cataracts and retinal
RT detachment: evidence for molecular heterogeneity in the Wagner
RT syndrome and the Stickler syndrome (arthro-ophthalmopathy).";
RL Am. J. Hum. Genet. 53:55-61(1993).
RN [20]
RX VARIANT SEMD CYS-840.
RA Tiller G.E.; Weis M.A.; Lachman R.S.; Cohn D.H.; Rimoin D.L.;
RA Eyre D.R.;
RT "A dominant mutation in the type II collagen gene (COL2A1) produces
RT spondyloepiphyseal dysplasia (SEMD), Strudwick type.";
RL Am. J. Hum. Genet. 53:A209-A209(1993).
RN [21]
RX VARIANT OSTEOARTHRITIS CYS-650.
RA MEDLINE=93282819; PubMed=8507190;
RA Holderbaum D.; Mäkelä C.J.; Moskowitz R.W.; Haqqi T.M.;
RT "Human cartilage from late stage familial osteoarthritis transcribes
RT type II collagen mRNA encoding a cysteine in position 519.";
RL Biochem. Biophys. Res. Commun. 192:1169-1174(1993).
RN [22]
RX VARIANT SEMD ARG-285.
RA MEDLINE=93252400; PubMed=8486375;

RA Vikkula M., Ritvaniemi P., Vuorio A.F., Kaitila I., Ala-Kokko L.,
 RA Peltonen L.,
 RA "A mutation in the amino-terminal end of the triple helix of type II
 RT collagen causing severe osteochondrodysplasia.",
 RL Genomics 16:282-285(1993).
 RN [23]
 RP VARIANT SEDC CYS-206.
 RX MEDLINE=94063862; PubMed=8244341;
 RA Williams C.J., Considine E.L., Knowlton R.G., Reginato A., Neumann G.,
 RA Harrison D., Buxton P., Jimenez S.A., Prockop D.J.;
 RT "Spondyloepiphyseal dysplasia and precocious osteoarthritis in a
 RT family with an Arg75->Cys mutation in the procollagen type II gene
 RT (COL2A1).",
 RL Hum. Genet. 92:499-505(1993).
 RN [24]
 RP VARIANT SEDC CYS-920.
 RX MEDLINE=93315508; PubMed=8325895;
 RA Chan D., Taylor T.K.F., Cole W.G.;
 RA "Characterization of an arginine 789 to cysteine substitution in
 RT alpha 1 (II) collagen chains of a patient with spondyloepiphyseal
 RT dysplasia.",
 RL J. Biol. Chem. 268:15238-15245(1993).
 RN [25]
 RP VARIANT SEDC SER-1128.
 RX MEDLINE=93140139; PubMed=8423604;
 RA Cole W.G., Hall R.K., Rogers J.G.;
 RA "The clinical features of spondyloepiphyseal dysplasia congenita
 RT resulting from the substitution of glycine 997 by serine in the alpha
 RT 1(II) chain of type II collagen.",
 RL J. Med. Genet. 30:27-35(1993).
 QY Query Match 72.8%; Score 862; DB 1; Length 1418;
 Best Local Similarity 74.0%; Pred. No. 1.4e-34;
 Matches 154; Conservative 11; Mismatches 43; Indels 0; Gaps 0;
 QY 1 GPPGPGPTGLPPGPGGSGFPCADGVAGKPGAGGSGPGAGKSPGAGRP 60
 DB 423 GARGEPGVGIGPPGEGKAPGNRFPQDGLAGKAGRGSGSLAGPKANGDFCRP 482
 QY 61 GEAGLPAGKJTGSPGSPGKGTGPPGAGQDGRPGPPGAGGAGVGMFGPKGAA 120
 DB 483 GEPGLGARGLTGRPDGAPQGVKVPFGAPGDEGPPGPPGQAGGQGVGMFGPKGAN 542
 QY 121 GEPGKAGRGVPGPVGAVGKAGGAGAGGPPGAGGAGGQGGAGSGFGGLPGPA 180
 DB 543 GEPGKAGRGVPGPVGAVGKAGGAGAGGPPGAGGAGGQGGAGSGFGGLPGPP 602
 QY 181 GPPGAGKPGQGVGDLGAPGSPGAG 208
 DB 603 GPPGSGKPGDQGVGEAGAGLVGPRG 630
 RESULT 8
 CA12 MOUSE STANDARD; PRT; 1459 AA.
 ID CA12 MOUSE
 AC P28481.
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Collagen alpha 1(II) chain precursor [Contains: Chondrocalcin].
 GN COL2A1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RX MEDLINE=91358489; PubMed=1885613;
 RA Metsaranta M., Toman D., de Crombrughe B., Vuorio E.;
 RT "Mouse type II collagen gene. Complete nucleotide sequence, exon
 RT structure, and alternative splicing.",
 RL J. Biol. Chem. 266:16862-16869(1991).
 RN [2]

RP SEQUENCE OF 1455-1459 FROM N.A.
 RX MEDLINE=91274355; PubMed=2054384;
 RA Metsaranta M., Toman D., de Crombrughe B., Vuorio E.;
 RT "Specific hybridization probes for mouse type I, II, III and IX
 RT collagen mRNAs.",
 RL Biochim. Biophys. Acta 1089:241-243(1991).
 CC -!- FUNCTION: Collagen type II is specific for cartilaginous tissues.
 CC -!- SUBUNIT: Homotrimers of alpha 1(II) chains.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=Long;
 CC IsoId=P28481-1; Sequence=Displayed;
 CC Name=Short;
 CC IsoId=P28481-2; Sequence=VSP 001139, VSP 001140;
 CC -!- PTM: Prolines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -!- SIMILARITY: Contains 1 VWF domain.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M65161; AAA68100.1; -;
 CC EMBL; X57982; AAA41047.1; -;
 CC MGD; MGI:88452; Col2a1.
 CC InterPro; IPR008161; Clg_helix.
 CC InterPro; IPR008160; Collagen.
 CC InterPro; IPR00885; Fib_collagen_C.
 CC InterPro; IPR001007; VWF_C.
 CC Pfam; PF01410; COLFI; 1.
 CC Pfam; PF01391; Collagen; 18.
 CC DR Pfam; PF00093; vwc; 1.
 CC DR ProDom; PD000007; Clg_helix; 5.
 CC DR ProDom; PD002078; Fib_collagen_C; 1.
 CC DR SMART; SM00038; COLFI; 1.
 CC DR SMART; SM00214; VMC; 1.
 CC DR PROSITE; PS01208; VWF_C; 1.
 CC DR PROSITE; PS0184; VWF_C; 1.
 CC DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 CC Glycoprotein; Collagen; Signal; Alternative splicing.
 FT SIGNAL 1 25 POTENTIAL
 FT PROPEP 26 153 AMINO-TERMINAL PROPEPTIDE
 FT (BY SIMILARITY).
 FT CHAIN 154 1213 COLLAGEN ALPHA 1(II) CHAIN.
 FT PROPEP 1214 1459 CHONDROCALCIN.
 FT DOMAIN 32 89 VWF.
 FT DOMAIN 173 1186 TRIPLE-HELICAL REGION.
 FT DOMAIN 1187 1213 NONHELICAL REGION (C-TERMINAL).
 FT VARSPLIC 29 29 Q -> R (in isoform Short).
 FT VARSPLIC 30 98 /FTId=VSP 001139.
 FT Missing (in isoform Short).
 FT /FTId=VSP 001140.
 SQ SEQUENCE 1459 AA; F6C84FA7C532E7F2 CRC64;
 Query Match 70.9%; Score 839; DB 1; Length 1459;
 Best Local Similarity 72.1%; Pred. No. 1.7e-33;
 Matches 150; Conservative 12; Mismatches 46; Indels 0; Gaps 0;
 QY 1 GPPGPGPTGLPPGPGGSGFPCADGVAGKPGAGGSGPGAGKSPGAGRP 60
 DB 464 GARGEPGAGPTGPPGEGKAPGNRFPQDGLAGKAGRGSGSLAGPKANGDFCRP 523
 QY 61 GEAGLPAGKJTGSPGSPGKGTGPPGAGQDGRPGPPGAGGAGVGMFGPKGAA 120
 DB 524 GEPGLGARGLTGRPDGAPQGVKVPFGAPGDEGPPGPPGQAGGQGVGMFGPKGAN 583
 QY 121 GEPGKAGRGVPGPVGAVGKAGGAGAGGPPGAGGAGGQGGAGSGFGGLPGPA 180
 DB 584 GEPGKAGRGVPGPVGAVGKAGGAGAGGPPGAGGAGGQGGAGSGFGGLPGPP 643

QY 181 GPPGAGKPGQGVPGDLGAPGSPGAG 208
 DB 644 GPPGEGKQGGDQIPGEAGAPGLVGRG 671

RESULT 9
 CALL_BOVIN
 ID CALL_BOVIN STANDARD; PRT; 779 AA.
 AC P02453;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Collagen alpha 1(I) chain (Fragments).
 GN COL1A1
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OC NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE OF 1-19.
 RX MEDLINE=72255334; PubMed=4115172;
 RA Rautenberg J., Timpl R., Furthmayr H.;
 RT "Structural characterization of N-terminal antigenic determinants in
 calf and human collagen.";
 RL Eur. J. Biochem. 27:231-237(1972).
 RN [2]
 RP SEQUENCE OF 20-145.
 RX MEDLINE=76022320; PubMed=1164916;
 RA Fietzek P.P., Kuehn K.;
 RT "The covalent structure of collagen: amino-acid sequence of the
 cyanogen-bromide peptides alpha-1-CB2, alpha-1-CB4 and alpha-1-CB5
 from calf-skin collagen.";
 RL Eur. J. Biochem. 52:77-82(1975).
 RN [3]
 RP SEQUENCE OF 146-294.
 RX MEDLINE=73049499; PubMed=4673951;
 RA Fietzek P.P., Wendt P., Kell I., Kuehn K.;
 RT "The covalent structure of collagen: amino acid sequence of alpha-1-
 CB3 from calf skin collagen.";
 RL FEBS Lett. 26:74-76(1972).
 RN [4]
 RP SEQUENCE OF 295-562.
 RX MEDLINE=74086118; PubMed=4359390;
 RA Fietzek P.P., Rexrodt F.W., Hopper K.E., Kuehn K.;
 RT "The covalent structure of collagen. 2. The amino-acid sequence of
 alpha-1-CB7 from calf-skin collagen.";
 RL Eur. J. Biochem. 38:396-400(1973).
 RN [5]
 RP SEQUENCE OF 563-675.
 RX MEDLINE=73042276; PubMed=4343808;
 RA Wendt P., Mark K.V.D., Rexrodt F., Kuehn K.;
 RT "The covalent structure of collagen. The amino-acid sequence of the
 112-residues. Amino-terminal part of peptide alpha-1-CB6 from calf-
 skin collagen.";
 RL Eur. J. Biochem. 30:169-183(1972).
 RN [6]
 RP SEQUENCE OF 676-751.
 RX MEDLINE=73042275; PubMed=4343807;
 RA Fietzek P.P., Rexrodt F.W., Wendt P., Stark M., Kuehn K.;
 RT "The covalent structure of collagen. Amino-acid sequence of peptide
 alpha-1-CB6-C2.";
 RL Eur. J. Biochem. 30:163-168(1972).
 CC -1- FUNCTION: Type I collagen is a member of group I collagen
 (fibrillar forming collagen).
 CC -1- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
 CC -1- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
 bones. In bones the fibrils are mineralized with calcium
 hydroxyapatite.
 CC -1- PMW: Proline residues at the third position of the tripeptide
 repeating unit (G-X-Y) are hydroxylated in some or all of the
 chains. Pro-726 is the only 3-hydroxypro and the only hydroxylated

CC proline in position X.
 CC -1- PMW: O-linked glycan consists of a Glc-Gal disaccharide bound to
 CC the oxygen atom of a post-translationally added hydroxyl group.
 CC -1- MISCELLANEOUS: The complete chain contains 1052 residues.
 DR PIR; A91193; CCB01S.
 DR InterPro; IPR008161; Clg_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR001007; WVF_C.
 DR ProDom; PD000007; Clg_helix; 1.
 DR PROSITE; PS01208; WVF_C_1; PARTIAL.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen; Pyrrolidone carboxylic acid.
 FT MOD_RES 1 1
 FT MOD_RES 9 9
 FT MOD_RES 103 103
 FT CARBOHYD 103 103
 FT MOD_RES 115 115
 FT MOD_RES 124 124
 FT MOD_CONS 145 146
 FT MOD_RES 274 274
 FT MOD_RES 346 346
 FT MOD_RES 424 424
 FT MOD_RES 496 496
 FT MOD_RES 658 658
 FT MOD_RES 670 670
 FT MOD_RES 726 726
 SQ SEQUENCE 779 AA; 70346 MW; E554A7FF084283D1 CRC64;
 Query Match 69.6%; Score 824; DB 1; Length 779;
 Best Local Similarity 72.0%; Pred. No. 5.8e-33;
 Matches 154; Conservative 4; Mismatches 50; Indels 6; Gaps 1;
 QY 1 GPPGPGPTGLPPPPGEGGPGS-----RGPPGADGVAGPKGAGRGSPGAPGKSP 54
 DB 29 GPPGAPGQGGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPG 88
 QY 55 GAGRPGGAGLPGAKLGTGSPGPGDGKTGPPGAGQDGRPPPPGPPGARGAGVMGFP 114
 DB 89 GARGLPGTAGLPGMKHGRGFSGLDCAKGDAGPAGPKGPGSGENGAPGQMGPRGLPGFP 148
 QY 115 GPKGAGGPGKAGRGVPPGAVGPGAGKDGAGAGQGGPGGAGGEGGPGAGSPGQFQ 174
 DB 149 GPKGAGGPGKAGRGVPPGAVGPGAGKDGAGAGQGGPGGAGGEGGPGAGSPGQFQ 208
 QY 175 GLPGGAGPFGAGKPGEGQGVPGDLGAPGSPGAG 208
 DB 209 GLPGGAGPFGAGKPGEGQGVPGDLGAPGSPGAG 242

RESULT 10
 CALL3_CHICK
 ID CALL3_CHICK STANDARD; PRT; 1262 AA.
 AC P12105; P79759; P79759; Q90794; Q92029;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Collagen alpha 1(III) chain precursor (Fragments).
 GN COL3A1.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus
 OC NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE OF 1-886 FROM N.A.
 RC TISSUE=Kidney;
 RX MEDLINE=94266842; PubMed=8206952;
 RA Nah H.-D., Niu Z., Adams S.L.;
 RT "An alternative transcript of the chick type III collagen gene that
 does not encode type III collagen.";
 RL J. Biol. Chem. 269:16443-16448(1994).
 RN [2]

SEQUENCE OF 29-96; 332-397; 431-484; 503-535 AND 869-976 FROM N.A.
 MEDLINE=84270696; PubMed=6547770;
 Yanada Y., Liou G., Mudryj M., Odici S., de Crombrughe B.;
 "Conservation of the sizes for one but not another class of exons in
 two chick collagen genes.";
 Nature 310:333-337(1984).
 [3]
 SEQUENCE OF 977-1262 FROM N.A.
 MEDLINE=83220816; PubMed=6856474;
 Yanada Y., Kuhn K., de Crombrughe B.;
 "A conserved nucleotide sequence, coding for a segment of the C-
 propeptide, is found at the same location in different collagen
 genes.";
 Nucleic Acids Res. 11:2733-2744(1983).
 CC -!- FUNCTION: Collagen type III occurs in most soft connective tissues
 along with type I collagen.
 CC -!- SUBUNIT: Trimers of identical alpha 1(III) chains. The chains are
 linked to each other by interchain disulfide bonds. Trimers are
 also cross-linked via hydroxylsines.
 CC -!- PTM: Prolines at the third position of the tripeptide repeating
 unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -!- SIMILARITY: Contains 1 VWFC domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 DR EMBL; U07973; AA83407.1; -
 DR EMBL; X00822; CAB52866.1; -
 DR EMBL; X00823; CAB52866.1; JOINED.
 DR EMBL; X00826; CAA25397.1; ALT_SEQ.
 DR EMBL; X00825; CAA25397.1; JOINED.
 DR EMBL; X00827; CAA25398.1; -
 DR EMBL; X00828; CAA25399.1; -
 DR EMBL; X00830; CAA25401.1; -
 DR EMBL; X00831; CAA25402.1; -
 DR EMBL; K02302; AAD15299.1; -
 DR EMBL; K02301; AAD15298.1; -
 DR EMBL; M36662; AAA18519.1; ALT_SEQ.
 DR PIR; A05269; A05269.
 DR PIR; I50694; I50694.
 DR InterPro; IPR008161; Clg_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR000885; Fib_collagen_C.
 DR InterPro; IPR002181; Fibrinogen_C.
 DR InterPro; IPR001007; VWF_C.
 DR ProDom; PD000007; Clg_helix; 1.
 DR ProDom; PD002078; Fib_collagen_C; 1.
 DR SMART; SM00038; COLF1; 1.
 DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS01208; VWC_1; 1.
 DR PROSITE; PS0184; VWC_2; 1.
 DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen; Signal.
 FT SIGNAL; 1
 FT PROPEP; 24 144
 FT CHAIN; 145 1003
 FT PROPEP; 1004 1262
 FT DOMAIN; 29 88
 FT DOMAIN; 145 164
 FT DOMAIN; 165 994
 FT DOMAIN; 995 1003
 FT NON_CONS; 886 887
 FT NON_CONS; 922 923
 FT DISULFID; 994 994
 FT INTERCHAIN (BY SIMILARITY).

FT DISULFID 995 995 INTERCHAIN (BY SIMILARITY).
 FT MOD_RES 262 262 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 283 283 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 859 859 HYDROXYLATION (BY SIMILARITY).
 FT CARBOHYD 1163 1163 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CONFLICT 96 96 E -> K (IN REF. 2).
 FT CONFLICT 1132 1132 F -> S (IN REF. 3).
 SQ SEQUENCE 1262 AA; 121249 MW; 96ABE7B2E9DEB43D CRC64;
 Query Match 67.1%; Score 794.5; DB 1; Length 1262;
 Best Local Similarity 65.9%; Pred. No. 2e-31;
 Matches 143; Conservative 13; Mismatches 52; Indels 9; Gaps 1;
 Qy 1 GPP-----GSPQPTGLPFPGERGPGSRGPGFADGAVGPKGPGAGERGSPGPK 51
 Db 458 GPPGEEGKRGANGEPGQNGVPTGGERGSPGRGLPGSNGLPGEKGPAGERGSPGPPG 517
 Qy 52 GSPEAGRPCEAGLPQAKGLTSPGSPGDKTGPAGQDGRPGPPGPPGARGQAGYM 111
 Db 518 GPAGDGGQGGGLFCMRLGLPIGSPGSDGKPGPGNGQGERGSPGPPGPPGPPG 577
 Qy 112 GFPPGKGAAGEPKAGRGVPPGPGAVGAGKDGAGAGGPPGPPGPPGPPGPPGPP 171
 Db 578 GFPPGKNEGAPGKNGERGPGPPGTPGAGKNGDVGLFPGPPGPPGPPGPPGPP 637
 Qy 172 GFQGLPQAPGPGCEAGKPGGQGVGGLGAPGSPGAG 208
 Db 638 GLOGLPQGPAGENGKPGEPGKGDIGGPPGPPGK 674
 RESULT 11
 CA25 HUMAN
 ID CA25 HUMAN STANDARD; PRT; 1496 AA.
 AC P05997;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Collagen alpha 2(V) chain precursor.
 GN COL5A2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE OF 1-463 FROM N.A.
 RX MEDLINE=89123368; PubMed=2914927;
 RA Woodbury D., Benson-Chanda V., Ramirez F.;
 "Amino-terminal propeptide of human pro-alpha 2(V) collagen conforms
 to the structural criteria of a fibrillar procollagen molecule.";
 J. Biol. Chem. 264:2735-2738(1989).
 RN [2]
 RP SEQUENCE OF 398-1496 FROM N.A.
 RX MEDLINE=87146331; PubMed=3029669;
 RA Weil D., Bernard M.P., Gargano S., Ramirez F.;
 "The pro alpha 2(V) collagen gene is evolutionarily related to the
 major fibrillar-forming collagens.";
 Nucleic Acids Res. 15:181-198(1987).
 RN [3]
 RP SEQUENCE OF 1227-1496 FROM N.A.
 RX MEDLINE=85289337; PubMed=2411731;
 RA Myers J.C., Loidl H.R., Seyer J.M., Dion A.S.;
 "Complete primary structure of the human alpha 2 type V procollagen
 COOH-terminal propeptide.";
 J. Biol. Chem. 260:11216-11222(1985).
 RN [4]
 RP SEQUENCE OF 1449-1496 FROM N.A.
 RX MEDLINE=89138450; PubMed=3224983;
 RA Tsipouras P., Schwartz R.C., Liddell A.C., Salkeld C.S., Weil D.,
 Ramirez F.;
 "Genetic distance of two fibrillar collagen loci, COL3A1 and COL5A2,
 located on the long arm of human chromosome 2.";
 Genomics 3:275-277(1988).
 RN [5]

RT chain of human type III procollagen. Differences in protein structure
 RT from type I procollagen and conservation of codon preferences.";
 RL Biochem. J. 260:509-516(1989).
 RP [2]
 RP SEQUENCE OF 149-1225 FROM N.A.
 RX MEDLINE=89386015; PubMed=2780304;
 RA Janeczko R.A., Ramirez F.;
 RT "Nucleotide and amino acid sequences of the entire human alpha 1
 RT (III) collagen.";
 RL Nucleic Acids Res. 17:6742-6742(1989).
 RN [3]
 RP SEQUENCE OF 168-398.
 RX MEDLINE=77134724; PubMed=557335;
 RA Seyer J.M., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of cyanogen
 RT bromide peptides from the amino-terminal segment of type III collagen
 RT of human liver.";
 RL Biochemistry 16:1158-1164(1977).
 RN [4]
 RP REVISIONS.
 RA Seyer J.M.;
 RL Submitted (DEC-1977) to the PIR data bank.
 RN [5]
 RP SEQUENCE OF 399-727.
 RX MEDLINE=7900343; PubMed=687591;
 RA Seyer J.M., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of five
 RT consecutive CNBr peptides from type III collagen of human liver.";
 RL Biochemistry 17:3404-3411(1978).
 RN [6]
 RP SEQUENCE OF 728-964.
 RX MEDLINE=80198282; PubMed=6246925;
 RA Seyer J.M., Mainardi C., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of alpha 1
 RT (III)-CB5 from type III collagen of human liver.";
 RL Biochemistry 19:1583-1589(1980).
 RN [7]
 RP SEQUENCE OF 950-1466 FROM N.A.
 RX MEDLINE=88189827; PubMed=3357782;
 RA Mankoo B.S., Dalgleish R.;
 RT "Human pro alpha 1(III) collagen: cDNA sequence for the 3' end.";
 RL Nucleic Acids Res. 16:2337-2337(1988).
 RN [8]
 RP REVISION TO 1184.
 RX MEDLINE=89098346; PubMed=3211760;
 RA Molyneux K., Dalgleish R.;
 RT "Human type III collagen 'variant' is a cDNA cloning artefact.";
 RL Nucleic Acids Res. 16:11833-11833(1988).
 RN [9]
 RP SEQUENCE OF 1065-1466 FROM N.A.
 RX MEDLINE=85087944; PubMed=6096827;
 RA Loidl H.R., Brinker J.M., May M., Pihlajaniemi T., Morrow S.,
 RA Rosenbloom J., Myers J.C.;
 RT "Molecular cloning and carboxyl-propeptide analysis of human type III
 RT procollagen.";
 RL Nucleic Acids Res. 12:9383-9394(1984).
 RN [10]
 RP SEQUENCE OF 965-1200.
 RX MEDLINE=81208339; PubMed=7016180;
 RA Seyer J.M., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of alpha
 RT 1(III)-CB9 from type III collagen of human liver.";
 RL Biochemistry 20:2621-2627(1981).
 RN [11]
 RP SEQUENCE OF 1176-1466 FROM N.A.
 RX MEDLINE=85157600; PubMed=2579949;
 RA Chu M.-L., Weil D., de Wet W.J., Bernard M.P., Sipola M., Ramirez F.;
 RT "Isolation of cDNA and genomic clones encoding human pro-alpha 1
 RT (III) collagen. Partial characterization of the 3' end region of the
 RT gene.";
 RL J. Biol. Chem. 260:4357-4363(1985).
 RN [12]
 RP SEQUENCE OF 1161-1200 FROM N.A.

RX MEDLINE=86187804; PubMed=3754462;
 RA Miskulin M., Dalgleish R., Kluever-Beckerman B., Rennard S.I.,
 RA Tolstoshev P., Brantly M., Crystal R.G.;
 RT "Human type III collagen gene expression is coordinately modulated
 RT with the type I collagen genes during fibroblast growth.";
 RL Biochemistry 25:1408-1413(1986).
 RN [13]
 RP SEQUENCE OF 1-170 FROM N.A.
 RX TISSUE=Placenta;
 RA MEDLINE=88303360; PubMed=3405773;
 RA Toman D., Ricca G., de Crombrughe B.;
 RT "Nucleotide sequence of a cDNA coding for the amino-terminal region
 RT of human prepro alpha 1(III) collagen.";
 RL Nucleic Acids Res. 16:7201-7201(1988).
 RN [14]
 RP SEQUENCE OF 1-176 FROM N.A.
 RX MEDLINE=89378752; PubMed=2777083;
 RA Benson-Chanda V., Su M.W., Weil D., Chu M.-L., Ramirez F.;
 RT "Cloning and analysis of the 5' portion of the human type-III
 RT procollagen gene (COL3A1).";
 RL Gene 78:255-265(1989).
 RN [15]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=97255959; PubMed=9101290;
 RA Kuivaniemi H., Tromp G., Prockop D.J.;
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
 RT associated collagen (type IX), and network-forming collagen (type X)
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
 RL Hum. Mutat. 9:300-315(1997).
 RN [16]
 RP VARIANT AORTIC ANEURYSM ARG-303, AND VARIANT THR-668.
 RX MEDLINE=93293988; PubMed=8514866;
 RA Tromp G., Wu Y., Prockop D.J., Madhathari S.L., Kleinert C.,
 RA Earley J.J., Zhuang J., Noergaard O., Darling R.C., Abbott W.M.,
 RA Cole C.W., Jaakkola P., Ryyanen M., Pearce W.H., Yao J.S.T., S.A.,
 RA Majamaa K., Smellens S.N., Gatalica Z., Ferrell R.E., Jimenez S.A.,
 RA Jackson C.E., Michels V.V., Kaye M., Kuivaniemi H.;
 RT "Sequencing of cDNA from 50 unrelated patients reveals that mutations
 RT in the triple-helical domain of type III procollagen are an
 RT infrequent cause of aortic aneurysms.";
 RL J. Clin. Invest. 91:2539-2545(1993).
 RN [17]
 RP VARIANT THR-698.
 RX MEDLINE=91045136; PubMed=2235526;
 RA Zafarullah K., Kleinert C., Tromp G., Kuivaniemi H., Kontusaari S.,
 RA Wu Y., Ganguly A., Prockop D.J.;
 RT "G to A polymorphism in exon 31 of the COL3A1 gene.";
 RL Nucleic Acids Res. 18:6180-6180(1990).
 RN [18]
 RP VARIANT AORTIC ANEURYSM ARG-786.
 RX MEDLINE=91056145; PubMed=2243125;
 RA Kontusaari S., Tromp G., Kuivaniemi H., Romanic A.M., Prockop D.J.;
 RT "A mutation in the gene for type III procollagen (COL3A1) in a family
 RT with aortic aneurysms.";
 RL J. Clin. Invest. 86:1465-1473(1990).
 RN [19]
 RP VARIANT EDS-IV ARG-828.
 RX MEDLINE=94016385; PubMed=8411057;
 RA Richards A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.;
 RT "The substitution of glycine 661 by arginine in type III collagen
 RT produces mutant molecules with different thermal stabilities and
 RT causes Ehlers-Danlos syndrome type IV.";
 RL J. Med. Genet. 30:690-693(1993).
 RN [20]
 RP VARIANT EDS-IV SER-957.
 RX MEDLINE=89109135; PubMed=2492273;
 RA Tromp G., Kuivaniemi H., Shikata H., Prockop D.J.;
 RT "A single base mutation that substitutes serine for glycine 790 of
 RT the alpha 1 (III) chain of type III procollagen exposes an arginine
 RT and causes Ehlers-Danlos syndrome IV.";
 RL J. Biol. Chem. 264:1349-1352(1989).
 RN [21]
 RP VARIANT EDS-IV VAL-960.

RA MEDLINE=95268429; PubMed=7749417;
RA Tromp G., de Paeppe A., Nuytinck L., Madhathari S.L., Kuivaniemi H.,
RT "Substitution of valine for glycine 793 in type III procollagen in
RT Ehlers-Danlos syndrome type IV.",
RL Hum. Mutat. 5:179-181(1995).
RL [22]
RP VARIANT EDS-IV GLU-1014.
RX MEDLINE=92316511; PubMed=1352273;
RA Richards A.J., Ward P.N., Narcisi P., Nicholls A.C., Lloyd J.C.,
RA Pope F.M.,
RT "A single base mutation in the gene for type III collagen (COL3A1)
RT converts glycine 847 to glutamic acid in a family with Ehlers-Danlos
RT syndrome type IV. An unaffected family member is mosaic for the
RT mutation.",
RL Hum. Genet. 89:414-418(1992).
RL [23]
RP VARIANT EDS-IV ASP-1050.
RX MEDLINE=90037070; PubMed=2808425;
RA Tromp G., Kuivaniemi H., Stollie C.A., Pope F.M., Prockop D.J.,
RT "Single base mutation in the type III procollagen gene that converts
RT the codon for glycine 883 to aspartate in a mild variant of
RT Ehlers-Danlos syndrome IV.",
RL J. Biol. Chem. 264:19313-19317(1989).
RL [24]
RP VARIANT EDS-IV VAL-1077.
RX MEDLINE=91374480; PubMed=1895316;
RA Richards A.J., Lloyd J.C., Ward P.N., de Paeppe A., Narcisi P.,
RA Pope F.M.,
RT "Characterisation of a glycine to valine substitution at amino acid
RT position 910 of the triple helical region of type III collagen in a
RT patient with Ehlers-Danlos syndrome type IV.",
RL J. Med. Genet. 28:459-463(1991).
RL [25]
RP VARIANT EDS-IV GLU-1173.
RX MEDLINE=93025543; PubMed=1357232;
RA Johnson P.H., Richards A.J., Pope F.M., Hopkinson D.A.,
Query Match 65.5%; Score 776; DB 1; Length 1466;
Best Local Similarity 66.3%; Pred. No. 1.6e-30;
Matches 138; Conservative 12; Mismatches 58; Indels 0; Gaps 0;
QY 1 GPPGEPGTGTPGPPGCGSGRGFGADGAGVAGPKGAPGAGSPGAGKSPGAGRGP 60
DB 468 GSPGEPGANGLPAGAGGKAGGAFGRGAPGPNIPGKGPAGGAPGAGPGAGFGRD 527
QY 61 GEAGLPAKGLTSGSPGSPGDKTGPAGQDGRPGPPGARGQAGVWFGPKGAA 120
DB 528 GVPGGPGVGRMGVSGFSGSGDKFGPGSQGSGRPGPGPGVWFGPKGND 587
QY 121 GEPKAGRGVPGPGVAGVPGAKGAGAGQPPGAPGAGRGEGGPGGPGGLPGPA 180
DB 588 GAPGKNGERGPGGPGGPGGPKNGKGTGPGGPGTGTGGDKGTGPPGQGLGPGTG 647
QY 181 GPPGEGAKPGGQGVGDLGAPGPGSPAG 208
DB 648 GPPGKNGKPGEPGPKGAGAGAGGKG 675
RESULT 13
CA21 RAT
ID CA21 RAT
AC P02466; 09R188; STANDARD; PRT; 1372 AA.
DT 21-JUL-1986 (Rel. 01, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 2(I) chain precursor.
GN COL1A2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.

RA Guenther D., Seibold S., Marx M.,
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RL [2]
RP SEQUENCE OF 86-98.
RC TISSUE=Skin;
RX MEDLINE=67162268; PubMed=5337886;
RA Kang A.H., Bornstein P., Piez K.A.,
RT "The amino acid sequence of peptides from the cross-linking region of
RT rat skin collagen.",
RL Biochemistry 6:788-795(1967).
RL [3]
RP SEQUENCE OF 99-102.
RC TISSUE=Skin;
RX MEDLINE=69206981; PubMed=5785232;
RA Fietzek P.P., Piez K.A.,
RT "Isolation and characterization of the cyanogen bromide peptides from
RT the alpha 2 chain of rat skin collagen.",
RL Biochemistry 8:2129-2133(1969).
RL [4]
RP SEQUENCE OF 102-144.
RC TISSUE=Skin;
RX MEDLINE=73049496; PubMed=4636752;
RA Fietzek P.P., Kell I., Kuehn K.,
RT "The covalent structure of collagen. Amino acid sequence of the N-
RT terminal region of alpha 2-CB4 from calf and rat skin collagen.",
RL FEBS Lett. 26:66-68(1972).
RL [5]
RP SEQUENCE OF 423-452.
RC TISSUE=Skin;
RX MEDLINE=71115216; PubMed=5544653;
RA Hieberger J.H., Kang A.H., Gross J.,
RT "Comparative studies on the amino acid sequence of the alpha 2-CB2
RT peptides from chick and rat skin collagens.",
RL Biochemistry 10:610-616(1971).
RL [6]
RP SEQUENCE OF 453-501.
RC TISSUE=Skin;
RX MEDLINE=75059250; PubMed=4435743;
RA Fietzek P.P., Kuehn K.,
RT "The covalent structure of collagen: amino acid sequence of the N-
RT terminal region of alpha2-CB3 from rat skin collagen and alpha2-CB3.5
RL from calf skin collagen.",
RL Hoppe-Seyler's Z. Physiol. Chem. 355:647-650(1974).
RL [7]
RP SEQUENCE OF 791-836.
RC TISSUE=Skin;
RX MEDLINE=74055004; PubMed=4763308;
RA Fietzek P.P., Kuehn K.,
RT "The covalent structure of collagen: amino acid sequence of the N-
RT terminal region of alpha 2-CB5 from rat skin collagen.",
RL FEBS Lett. 36:289-291(1973).
RL [8]
RP ORDER OF CNBR PEPTIDES.
RX MEDLINE=70181852; PubMed=5443712;
RA Vuust J., Lane J.M., Fietzek P.P., Miller E.J., Piez K.A.,
RT "The order of the CNBR peptides from the alpha 2 chain of collagen.",
RL Biochem. Biophys. Res. Commun. 38:703-708(1970).
RL CC -!- FUNCTION: Type I collagen is a member of group I collagen
CC (fibrillar forming collagen).
CC CC -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC CC -!- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
CC bones. In bones the fibrils are mineralized with calcium
CC hydroxyapatite.
CC CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

Search completed: September 24, 2004, 11:07:23
Job time : 7.27357 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2004, 11:06:56 ; Search time 8.04914 Seconds
(without alignments)
2497.663 Million cell updates/sec

Title: US-10-658-989a-1

Perfect score: 1184

Sequence: 1 GPPGEPGTGLPQPPGSRG.....GEGVPGDLGAPGSPAGG 209

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78:**

1: PIR1.*

2: PIR2.*

3: PIR3.*

4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1149	97.0	1464	1 CGHU1S	collagen alpha 1(I)
2	1118	94.4	671	1 CGRT1S	collagen alpha 1(I)
3	1116	94.3	1453	2 S21626	collagen alpha 1(I)
4	1067	90.1	1042	1 CGCH1S	collagen alpha 1(I)
5	862	72.8	1487	1 CGHU6C	collagen alpha 1(I)
6	860	72.6	1418	2 T45467	collagen alpha 1(I)
7	850	71.8	1486	1 B40333	collagen alpha 1(I)
8	847	71.5	1492	2 A40333	collagen alpha 1(I)
9	839	70.9	1419	2 A41182	collagen alpha 1(I)
10	839	70.9	1487	2 B41182	collagen alpha 1(I)
11	824	69.6	779	1 CGB01S	collagen alpha 1(I)
12	815	68.8	673	1 CGB06C	collagen alpha 1(I)
13	794.5	67.1	886	2 I50694	collagen alpha 1(I)
14	786	66.4	1496	1 CGHU2V	collagen alpha 2(V)
15	776	65.5	1466	1 CGHU7L	collagen alpha 1(I)
16	765	64.6	1497	2 I49607	procollagen type V
17	763	64.4	1464	2 S59856	collagen alpha 1(I)
18	761.5	64.3	1049	1 CGB07S	collagen alpha 1(I)
19	760	64.2	1373	1 A43291	collagen alpha 2(I)
20	744	62.8	1366	1 CGB07S	collagen alpha 2(I)
21	629.5	53.2	1027	2 S28774	collagen alpha 1(V)
22	607	51.3	1838	1 CGHU1V	collagen alpha 1(I)
23	602.5	50.9	310	2 I50696	collagen alpha 2(I)
24	600.5	50.7	1414	1 S23809	collagen alpha 1(I)
25	599	50.6	1843	2 S18803	collagen alpha 1(V)
26	590.5	49.9	632	2 S42731	collagen alpha 1(C)
27	589	49.7	964	1 CGCH2S	collagen alpha 2(I)
28	577	48.7	1806	1 CGHU1E	collagen alpha 1(X)
29	576.5	48.7	730	2 A36226	collagen alpha 1 C

RESULT 1

CGHU1S

collagen alpha 1(I) chain precursor - human

N;Alternate names: procollagen alpha 1(I) chain

C;Species: Homo sapiens (man)

C;Date: 12-Aug-1981 #sequence, revision 04-Oct-1996 #text_change 31-Dec-2000

C;Accession: I60114; S01143; A93335; I55254; A39943; I55237; A35233; S09400; B90567; S11289; A29439; I53466; A02852; I37247

R;D'Alessio, M.; Bernard, M.; Pretorius, P.J.; de Wet, W.; Ramirez, F.; Pretorius, P.J.

Gene 67, 105-115, 1988

A;Title: Complete nucleotide sequence of the region encompassing the first twenty-five ex

A;Reference number: I60114; MUID:88329734; PMID:2843432

A;Accession: I60114

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-369, 'L', 371-589 <DAL>

A;Cross-references: GB:M20789; NID:g179593; PIDN:AAB59373.1; PID:g179594

R;Tromp, G.; Kuhlmann, H.; Stacey, A.; Shikata, H.; Baldwin, C.T.; Jaenisch, R.; Prock

Biochem. J. 253, 919-922, 1988

A;Title: Structure of a full-length cDNA clone for the prepro-alpha-1(I) chain of human t

A;Reference number: S01143; MUID:89025644; PMID:3178743

A;Accession: S01143

A;Molecule type: mRNA

A;Residues: 1-472 <TRO>

A;Cross-references: EMBL:X07884; NID:g30015; PIDN:CAA30731.1; PID:g30016; GB:M36546; NID

A;Note: Submitted to the EMBL/GenBank/DBJ databases by Prockop, D.J., 13-JUN-1988

R;Chu, M.L.; de Wet, W.; Bernard, M.; Ding, J.F.; Morabito, M.; Myers, J.; Williams, C.;

Nature 310, 337-340, 1984

A;Title: Human proalpha1(I) collagen gene structure reveals evolutionary conservation of

A;Reference number: A93335; MUID:84270697; PMID:6462220

A;Accession: A93335

A;Molecule type: DNA

A;Residues: 1-58, 'Q', 60-181 <CHU>

A;Cross-references: EMBL:X00820; NID:g35657; PIDN:CAA25394.1; PID:g35658

R;Rossouw, C.M.S.; Vergeer, W.P.; du Plooy, S.J.; Bernard, M.P.; Ramirez, F.; de Wet, W.;

J. Biol. Chem. 262, 15151-15157, 1987

A;Title: DNA sequences in the first intron of the human pro-alpha 1(I) collagen gene enh

A;Reference number: I55254; MUID:88033098; PMID:2822714

A;Accession: I55254

A;Status: translation not shown; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-45 <ROS>

A;Cross-references: GB:J02829; NID:g180387; PIDN:AAA51993.1; PID:g180388

R;Bornstein, P.; McKay, J.; Morishima, J.K.; Devarayalu, S.; Gelinas, R.E.

Proc. Natl. Acad. Sci. U.S.A. 84, 8869-8873, 1987

A;Title: Regulatory elements in the first intron contribute to transcriptional control of

A;Reference number: A39943; MUID:98097389; PMID:3480516

A;Accession: A39943

A;Molecule type: DNA

A;Residues: 1-34 <BOR>

A;Cross-references: GB:J03559; NID:g180876; PIDN:AAA52052.1; PID:g553238

R;Chu, M.L.; de Wet, W.; Bernard, M.; Ramirez, F.

J. Biol. Chem. 260, 2315-2320, 1985
 A>Title: Fine structural analysis of the human pro-alpha 1 (I) collagen gene. Promoter s
 A;Reference number: I55237; MUID:85130970; PMID:2857713
 A;Accession: I55237
 A;Molecule type: mRNA
 A;Status: translation not shown; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-34 <CH2>
 A;Cross-references: GB:M0627; NID:g180383; PIDN:AA51992.1; PID:g553226
 R;Wittz, M.K.; Keene, D.R.; Hori, H.; Glanville, R.W.; Steinmann, B.; Rao, V.H.; Hollist
 J. Biol. Chem. 265, 6312-6317, 1990
 A>Title: In vivo and in vitro noncovalent association of excised alpha1(I) amino-termina
 rome, type VII.
 A;Reference number: A35233; MUID:90202908; PMID:2318855
 A;Accession: A35233
 A;Molecule type: protein
 A;Residues: 33-52 <WR>
 A;Note: this propeptide fragment remained non-covalently bound to a defective, uncleaved
 R;Well, D.; d'Alessio, M.; Ramirez, F.; de Wet, W.; Cole, W.G.; Chan, D.; Bateman, J.F.
 EMBO J. 8, 1705-1710, 1989
 A>Title: A base substitution in the exon of a collagen gene causes alternative splicing
 A;Reference number: S09400; MUID:89356643; PMID:2767050
 A;Accession: S09400
 A;Molecule type: mRNA
 A;Residues: 156-183 <WEI>
 R;Click, E.M.; Bornstein, P.
 Biochemistry 9, 4699-4706, 1970
 A>Title: Isolation and characterization of the cyanogen bromide peptides from the alpha
 A;Reference number: A90567; MUID:71038625; PMID:5529814
 A;Contents: CNBr0-1, CNBr2, CNBr4, CNBr5
 A;Accession: B90567
 A;Molecule type: protein
 A;Residues: 163-198, 'Z', 200-201, 'Z', 203-206, 'Z', 208-209, 'Z', 211-228, 'B', 230, 'BB', 233, 'Z'
 A;Experimental source: skin
 A;Note: evidence for 170-allysine
 R;Baetge, B.; Notbohm, H.; Diebold, J.; Lehmann, H.; Bodo, M.; Deutzmann, R.; Mueller, F.
 Eur. J. Biochem. 192, 153-159, 1990
 A>Title: A critical crosslink region in human-bone-derived collagen type I. Specific cle
 A;Reference number: S11372; MUID:90382436; PMID:2169412
 A;Accession: S11372
 A;Molecule type: protein
 A;Residues: 175-187, 274-287, 'P', 289 <BAE>
 A;Note: sequence of collagen alpha 1(S)(I) isolated from bone after pepsin digestion
 R;Deak, S.B.; Scholz, P.M.; Amenta, P.S.; Constantinou, C.D.; Levi-Minzi, S.A.; Gonzalez
 J. Biol. Chem. 266, 21827-21832, 1991
 A>Title: The substitution of arginine for glycine 85 of the alpha 1(I) procollagen chain
 A;Reference number: I55342; MUID:92042092; PMID:1718984
 A;Accession: I55342
 A;Status: translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 258-268, 1347-1357 <DEA>
 A;Cross-references: GB:S67495; NID:g239007; PIDN:AA20350.1; PID:g239008
 R;Morgan, P.H.; Jacobs, H.G.; Segrest, J.P.; Cunningham, L.W.
 J. Biol. Chem. 245, 5042-5048, 1970
 A>Title: Comparative study of glycopeptides derived from selected vertebrate collagens.
 A;Reference number: A92069; MUID:71001508; PMID:4319110
 A;Accession: A92069
 A;Molecule type: protein
 A;Residues: 263-268 <MOR>
 A;Experimental source: skin
 A;Note: attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose to 5-hydroxylysine
 R;Labhard, M.E.; Hollister, D.W.
 Matrix 10, 124-130, 1990
 A>Title: Segmental amplification of the entire helical and telopeptide regions of the ch
 A;Reference number: S15989; MUID:90326017; PMID:2374517
 A;Accession: S15989
 A;Molecule type: mRNA
 A;Residues: 281-302, 402-420, 823-843, 925-944, 1026-1045, 1143-1162 <LAB>
 R;Wittz, M.K.; Rao, V.H.; Glanville, R.W.; Labhard, M.E.; Pretorius, P.J.; de Vries, W.N.
 Connect. Tissue Res. 29, 1-11, 1993
 A>Title: A cysteine for glycine substitution at position 175 in an alpha 1 (I) chain of
 A;Reference number: I52905; MUID:93339042; PMID:8339541

A;Accession: I52905
 A;Status: translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 342-352, 'C', 354-359 <W12>
 A;Cross-references: GB:S64717; NID:g408195; PIDN:AA27677.1; PID:g408196
 A;Note: mutant sequence from patient with osteogenesis imperfecta
 R;Bernard, M.P.; Chu, M.L.; Myers, J.C.; Ramirez, F.; Eikenberry, E.F.; Prockop, D.J.
 Biochemistry 22, 5213-5223, 1983
 A>Title: Nucleotide sequences of complementary deoxyribonucleic acids for the proalpha1
 A;Reference number: A90476; MUID:84080385; PMID:6689127
 A;Accession: A90476
 A;Molecule type: mRNA
 A;Residues: 425-1250, 'X', 1252-1328, 'S', 1330-1390, 'X', 1392-1464 <BER>
 A;Cross-references: GB:K01228; NID:g180391; PIDN:AA51995.1; PID:g180392
 A;Note: sequence partially completed for missing nucleotides by A29439
 R;Chu, M.L.; Gargiulo, V.; Williams, C.J.; Ramirez, F.
 J. Biol. Chem. 260, 691-694, 1985
 A>Title: Multixon deletion in an osteogenesis imperfecta variant with increased type III
 A;Reference number: A22161; MUID:85104934; PMID:2981843
 A;Accession: A22161
 A;Molecule type: DNA
 A;Residues: 472-594, 'R', 596-607 <CH3>
 A;Cross-references: GB:K03178; GB:K03179; NID:g179612; NID:g179613; PIDN:AA51847.1; PID:
 A;Note: the authors translated the codon CGT for residue 595 as Pro
 R;Wallis, G.A.; Starman, B.J.; Zinn, A.B.; Byers, P.H.
 Am. J. Hum. Genet. 46, 1034-1040, 1990
 A>Title: Variable expression of osteogenesis imperfecta in a nuclear family is explained
 A;Reference number: A35336; MUID:90252792; PMID:2339700
 A;Accession: A35336
 A;Molecule type: mRNA
 A;Residues: 710-720, 'E', 722-737, 'E', 739-745 <WAL>
 A;Note: the authors translated the codons CAG for 721 and CGT for 738 as Glu
 R;Forlino, A.; Zolezzi, F.; Valli, M.; Pignatti, P.F.; Cetta, G.; Brunelli, P.C.; Mottes,
 Hum. Mol. Genet. 3, 2201-2206, 1994
 A>Title: Severe (type III) osteogenesis imperfecta due to glycine substitutions in the c
 A;Reference number: I54365; MUID:95187161; PMID:7881420
 A;Accession: I54365
 A;Status: translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 746-766, 'S', 768-781 <FOR>
 A;Cross-references: GB:L47667; NID:g1009093; PIDN:AA59576.1; PID:g1009094
 R;Chessler, S.D.; Wallis, G.A.; Byers, P.H.
 J. Biol. Chem. 268, 18218-18225, 1993
 A>Title: Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I) chain of ty
 A;Reference number: A47426; MUID:93352646; PMID:8349697
 A;Accession: A47426
 A;Molecule type: mRNA
 A;Residues: 1179-1276, 'H', 1278-1336, 1339-1387, 'R', 1389-1464 <CHE>
 A;Cross-references: GB:S64596; NID:g407589; PIDN:AA27856.1; PID:g407590
 A;Note: sequence extracted from NCBI backbone (NCBIN:136444, NCBI:P:136445)
 A;Note: does not represent an experimentally determined sequence but three different muta
 A;Accession: B47426
 A;Molecule type: mRNA
 A;Residues: 1179-1464 <CH4>
 A;Experimental source: normal dermal fibroblast culture
 A;Accession: C47426
 A;Molecule type: mRNA
 A;Residues: 1179-1276, 'H', 1278-1464 <CH5>
 A;Experimental source: fetal cell 86-237
 A;Accession: D47426
 A;Molecule type: mRNA
 A;Residues: 1179-1336, 1339-1464 <CH6>
 A;Experimental source: fetal cell 86-146
 A;Accession: E47426
 A;Molecule type: mRNA
 A;Residues: 1179-1387, 'R', 1389-1464 <CH7>
 A;Experimental source: fetal cell 88-251
 R;Cohn, D.H.; Apone, S.; Eyre, D.R.; Starman, B.J.; Andreassen, P.; Charbonneau, H.; Nicl
 J. Biol. Chem. 263, 14605-14607, 1988
 A>Title: Substitution of Cysteine for Glycine within the Carboxyl-terminal Telopeptide of
 A;Reference number: I55269; MUID:89008319; PMID:3170557
 A;Accession: I55269

A;Experimental source: skin
R;Butler, W.T.
Biochemistry 9, 44-50, 1970
A;Title: Chemical studies on the cyanogen bromide peptides of rat skin collagen. The cov
A;Reference number: A90566; MUID:70085124; PMID:5411206
A;Contents: CNBr5
A;Accession: A90566
A;Molecule type: protein
A;Residues: 103-139 <BU2>
A;Experimental source: skin
R;Balian, G.; Click, E.M.; Bornstein, P.
Biochemistry 10, 4470-4478, 1971
A;Title: Structure of rat skin collagen alphas 1(I). Amino acid sequence of the hydroxyl
A;Reference number: A90357; MUID:72136131; PMID:4335087
A;Contents: CNBr8
A;Accession: A90357
A;Molecule type: protein
A;Residues: 140-238 <BA1>
A;Experimental source: skin
R;Balian, G.; Click, E.M.; Hermanson, M.A.; Bornstein, P.
Biochemistry 11, 3798-3806, 1972
A;Title: Structure of rat skin collagen alphas 1(I). Amino acid sequence of the hydroxyl
A;Reference number: A90362; MUID:73008942; PMID:4342027
A;Contents: CNBr8
A;Accession: A90362
A;Molecule type: protein
A;Residues: 239-418 <BA2>
A;Experimental source: skin
R;Butler, W.T.; Underwood, S.P.; Finch Jr., J.E.
Biochemistry 13, 2945-2953, 1974
A;Title: Chemical studies on the cyanogen bromide peptides of rat skin collagen. Amino ac
A;Reference number: A90379; MUID:74271984; PMID:4366532
A;Contents: CNBr3
A;Accession: A90379
A;Molecule type: protein
A;Residues: 419-567 <BU3>
A;Experimental source: skin
R;Stoltz, M.; Timpl, R.; Furthmayr, H.; Kuehn, K.
Eur. J. Biochem. 37, 287-294, 1973
A;Title: Structural and immunogenic properties of a major antigenic determinant in neutr
A;Reference number: A91209; MUID:74011954; PMID:4126850
A;Contents: CNBr6
A;Accession: A91209
A;Molecule type: protein
A;Residues: 568-651 <ST1>
A;Experimental source: skin
A;Note: this region probably corresponds to positions 949-1032 of the alpha 1(I) chain
A;Note: the major antigenic determinant (of neutral salt-extracted rat skin collagen) inv
R;Stoltz, M.; Timpl, R.; Kuehn, K.
FEBS Lett. 26, 61-65, 1972
A;Title: Non-helical regions in rat collagen alpha 1-chain.
A;Reference number: A91385; MUID:73049495; PMID:4636751
A;Contents: CNBr6
A;Accession: A91385
A;Molecule type: protein
A;Residues: 651-671 <ST2>
A;Experimental source: skin
A;Note: the composition of peptides comprising residues 1-9 and 1-19 confirms the sequen
A;Note: this region (residues 651-671 above) probably corresponds to positions 1032-1052
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (o
ed and subsequently O-glycosylated.
C;Comment: The order of the nine CNBr peptides in the alpha 1(I) chain of rat skin colla
C;Comment: The complete chain contains 1052 residues.
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: blocked amino end; coiled coil; extracellular matrix; glycoprotein; hydroxyl
F;1/Modified site: blocked amino end (Glx) (probably pyrrolidone carboxylic acid) #statu
F;9/Modified site: allysine (Lys) #status experimental
F;103/424,547/Binding site: carboxylate (Lys) (covalent) #status experimental
F;103/Modified site: 5-hydroxylysine (Lys) #status experimental
F;424,547/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental

Matches 198; Conservative 5; Mismatches 5; Indels 12; Gaps 1;

QY 1 GPP-----GEPGTGLPQPPGGRGPGSRGPPGADGVAGPKGPPAGERGSPGPA 48
 DB 296 GPPGAGGKRGAGGEPGGLPQPPGGRGPGSRGPPGADGVAGPKGPPAGERGSPGPA 355
 QY 49 GPKGSPGAGRPGGAGLPGKAGLGTGSPGSPGDPGKTGPPGAGODGRPPGPPGARGQA 108
 DB 356 GPKGSPGAGRPGGAGLPGKAGLGTGSPGSPGDPGKTGPPGAGZBGRPPGPPGARGQA 415
 QY 109 GVMGPPGKGAABGPKGKAGRGVPGPPGAVGPAKGDGAGAGQPPGPPGAGERGEGGPA 169
 DB 416 GVMGPPGKGAABGPKGKAGRGVPGPPGAVGPAKGDGAGAGQPPGPPGAGERGEGGPA 475
 QY 169 GSPGQGLPGPAGPPGAGKPGEGQVPGDLGAPGSPGAG 208
 DB 476 GSPGQGLPGPAGPPGAGKPGEGQVPGDLGAPGSPGAG 515

RESULT 3
 S21626
 collagen alpha 1(I) chain precursor - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 13-Jan-1995 #sequence revision 25-Apr-1997 #text change 13-Aug-1999
 C:Accession: S57243; S16374; A23982; I49559; S39789; I48300; S21626
 R:Li, S.W.; Khillan, J.; Prockop, D.J.
 Matrix Biol. 14, 593-595, 1994
 A>Title: The complete cDNA coding sequence for the mouse pro-alpha-1(I) chain of type I
 A:Reference number: S57243
 A:Accession: S57243
 A:Molecule type: mRNA
 A:Residues: 1-1453 <LIS>
 A:Cross-references: EMBL:U08020; NID:9470673; PIDN:AAA88912.1; PID:G470674
 R:Metsaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.
 Biochim. Biophys. Acta 1089, 241-243, 1991
 A>Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.
 A:Reference number: S16176; MUID:91274355; PMID:2054384
 A:Accession: S16374
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1442-1453 <MET>
 A:Cross-references: EMBL:X57981; NID:G50484; PIDN:CAA1046.1; PID:G50485
 R:French, B.T.; Lee, W.H.; Maul, G.G.
 Gene 39, 311-312, 1985
 A>Title: Nucleotide sequence of a cDNA clone for mouse proalpha1(I) collagen protein.
 A:Reference number: A23982; MUID:86137403; PMID:3841523
 A:Accession: A23982
 A:Molecule type: mRNA
 A:Residues: 518-1128 <FRE>
 A:Cross-references: GB:M14423; NID:G192261; PIDN:AAA37333.1; PID:G192262
 R:Monson, J.M.; Friedman, J.; McCarthy, B.J.
 Mol. Cell. Biol. 2, 1362-1371, 1982
 A>Title: DNA sequence analysis of a mouse pro-alpha-1(I) procollagen gene: Evidence for
 A:Reference number: I49559; MUID:83141374; PMID:6298597
 A:Accession: I49559
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 735-1130 <RES>
 A:Cross-references: GB:M17491; NID:G192263; PIDN:AAA37334.1; PID:G192264
 R:Harbers, K.; Kuehn, H.; Jaenisch, R.
 Proc. Natl. Acad. Sci. U.S.A. 81, 1504-1508, 1984
 A>Title: Insertion of retrovirus into the first intron of alpha1(I) collagen gene leads
 A:Reference number: I49557; MUID:84170331; PMID:6324198
 A:Accession: I49557
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-25 <RES>
 A:Cross-references: GB:K01688; NID:G192246; PIDN:AAA37330.1; PID:G553881
 R:Fenton, S.P.; Lamande, S.R.; Hannagan, M.; Stacey, A.; Jaenisch, R.; Bateman, J.F.
 Biochim. Biophys. Acta 1216, 469-474, 1993
 A>Title: Genomic sequence of mouse COL1A1 encoding the collagen propeptides.
 A:Reference number: S39789; MUID:94092741; PMID:8268229
 A:Accession: S39789

A:Molecule type: DNA
 A:Residues: 1-80, 'E', 82-105, 'D', 107-185, 1031-1201, 'G', 1203-1218, 'E', 1220-1221, 'T', 1223-1224, 'R';
 R:Rhodes, K.; Rippe, R.A.; Umezawa, A.; Nehls, M.; Brenner, D.A.; Breindl, M.
 Mol. Cell. Biol. 14, 5950-5960, 1994
 A>Title: DNA methylation represses the murine alpha 1(I) collagen promoter by an indirect
 A:Reference number: I48300; MUID:94344105; PMID:8065328
 A:Accession: I48300
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-80, 'E', 82-105, 'D', 107-147 <REF>
 A:Cross-references: EMBL:X54876; NID:G50486; PIDN:CAA38657.1; PID:G50487
 C:Genetics: COL1A1
 A:Introns: 770/3; 788/3; 806/3; 842/3; 860/3; 878/3; 932/3; 968/3; 1004/3; 1022/3; 1058/3;
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 C:Keywords: Coiled coil; extracellular matrix; glycoprotein; heterotrimer; triple helix
 F;1-22/Domain: signal sequence #status predicted <SIG>
 F;23-151/Domain: amino-terminal propeptide #status predicted <PRO>
 F;30-89/Domain: von Willebrand factor type C repeat homology <VWC>
 F;152-1453/Product: collagen alpha 1(I) chain #status predicted <MAT>
 F;1224-1453/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 94.3%; Score 1116; DB 2; Length 1453;
 Best Local Similarity 90.5%; Pred. No. 4.5e-55;
 Matches 199; Conservative 4; Mismatches 5; Indels 12; Gaps 1;
 QY 1 GPP-----GEPGTGLPQPPGGRGPGSRGPPGADGVAGPKGPPAGERGSPGPA 48
 DB 447 GPPGAGGKRGAGGEPGGLPQPPGGRGPGSRGPPGADGVAGPKGPPAGERGSPGPA 506
 QY 49 GPKGSPGAGRPGGAGLPGKAGLGTGSPGSPGDPGKTGPPGAGODGRPPGPPGARGQA 108
 DB 507 GPKGSPGAGRPGGAGLPGKAGLGTGSPGSPGDPGKTGPPGAGZBGRPPGPPGARGQA 566
 QY 109 GVMGPPGKGAABGPKGKAGRGVPGPPGAVGPAKGDGAGAGQPPGPPGAGERGEGGPA 168
 DB 567 GVMGPPGKGAABGPKGKAGRGVPGPPGAVGPAKGDGAGAGQPPGPPGAGERGEGGPA 626
 QY 169 GSPGQGLPGPAGPPGAGKPGEGQVPGDLGAPGSPGAG 208
 DB 627 GSPGQGLPGPAGPPGAGKPGEGQVPGDLGAPGSPGAG 666

RESULT 4
 CGCH15
 collagen alpha 1(I) chain - chicken (tentative sequence) (fragments)
 C:Species: Gallus gallus (chicken)
 C>Date: 12-Aug-1981 #sequence revision 06-Jul-1982 #text change 31-Mar-2000
 C:Accession: A90458; A90181; A02857
 R:Highberger, J.H.; Corbett, C.; Dixit, S.N.; Yu, W.; Seyer, J.M.; Kang, A.H.; Gross, J.
 Biochemistry 21, 2048-2055, 1982
 A>Title: Amino acid sequence of chick skin collagen alpha 1(I)-C88 and the complete primary
 A:Reference number: A90458; MUID:82231995; PMID:7093229
 A:Accession: A90458
 A:Molecule type: protein
 A:Residues: 1-1036 <HIG>
 A:Experimental source: skin
 A>Note: This is the latest in a series of papers from these workers elucidating the sequence
 R:Evye, D.R.; Glimcher, M.J.
 Biochem. Biophys. Res. Commun. 48, 720-726, 1972
 A>Title: Evidence for a previously undetected sequence at the carboxyterminus of the alpha1
 A:Reference number: A90181; MUID:72243016; PMID:5047697
 A:Accession: A90181
 A:Molecule type: protein
 A:Residues: 1037-1042 <EYR>
 A:Experimental source: skin
 A>Note: residues 1037-1042 above correspond to the carboxyl end of the protein
 C:Comment: lysines at positions 103, 700, 934, and 946 above may be hydroxylated in some
 C:Comment: Most of the prolines at the third position of the tripeptide repeating unit (C
 C:Comment: Pro-1002 is the only 3-hydroxyproline and the only hydroxylated proline in pos
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 C:Keywords: coiled coil; extracellular matrix; glycoprotein; pyroglutamic acid; trimer; t
 F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match 90.1%; Score 1067; DB 1; Length 1042;
Best Local Similarity 86.4%; Pred.No.1.e-52;
Matches 190; Conservative 4; Mismatches 14; Indels 12; Gaps 1;

QY 1 GPSP-----GEPGPTGLPFPGBRGSGRGGFFCGADGVAGPKGPAGRGSPGPA 48
DB |
296 GPFGPAGESKRGCAEGPEPGAGLPGFAGERGAFSGRGGFFCGADGIAGPKGPPGERGSPGAV 355

QY 49 GPXGSGCEAGRCPGEAGLPGAKGLTGSFGSPGPPDGKTGTPPGPACQDGRPPGFPPGARGQA 108
DB 356 GPXGSPCEAGRCPGEAGLPGAKGLTGSFGSPGPPDGKTGTPPGPACQDGRPPGPAGPPGARGQA 415

QY 109 GVNGFFGPKGAACEPGKAGERGVPFGPVGAPVGPAGKDKGEACAQGPPGPPGAPGABRRGQQGPA 168
DB 416 GVNGFFGPKGAACEPGKAGERGVPFGPVGAPVGPAGKDKGEACAQGPPGPTGAPGERGQQGPA 475

QY 169 GSFGFGCLPGPAGPPGAEAGKPGGQGVPGDLGAPGPGSPAG 208
DB 476 GAPFGFGLPGPAGPPGAEAGKPGGQGVPGFNAGAPGAG 515

RESULT 5
CGHUC6C
collagen alpha 1(II) chain precursor [validated] - human
N;Alternate names: procollagen alpha 1(II) chain
N;Contains: chondrocalcin; collagen alpha 1(II) chain precursor splice form 1; collagen
C;Species: Homo sapiens (man)
C;Date: 28-May-1986 #sequence revision 01-Sep-1995 #text change 08-Dec-2000
C;Accession: A38513; S06715; S24270; A24828; S06496; A35428; A30147; A33116; S64674; S63
7250; I37251; I37252; I37253; I37254; I35338; I59335; I61910
R;Ryan, M.C.; Sisraski, M.; Sandell, L.J.
Genomics 8; 41-48, 1990
A;Title: The human type II procollagen gene: identification of an additional protein-coo
A;Reference number: A38513; MUID: 91184811; PMID: 2081599
A;Accession: A38513
A;Molecule type: DNA
A;Residues: 1-103 <RYA>
A;Cross-references: GB:M60299; NID:g180883; PIDN:AAA73873.1; PID:g180884
R;Su, M.W.; Lee, B.; Ramirez, F.; Machada, M.; Horton, W.
Nucleic Acids Res. 17, 9473, 1989
A;Title: Nucleotide sequence of the full length cDNA encoding for human type II procolla
A;Reference number: S06715; MUID: 90067946; PMID: 2597267
A;Accession: S06715
A;Molecule type: mRNA
A;Residues: 1-28 'R', 99-1487 <SU2>
A;Cross-references: EMBL:X16468; NID:g29515; PIDN:CAA34488.1; PID:g29516
A;Note: alternative splice form 1
R;Viikula, M.; Matsuranta, M.; Syvaenen, A.C.; Ala-Kokko, L.; Vuorio, E.; Peltonen, L.
Biochem. J. 285, 287-294, 1992
A;Title: Structural analysis of the regulatory elements of the type-II procollagen gene.
A;Reference number: S24270; MUID: 92344585; PMID: 1637314
A;Accession: S24270
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-28 <VIK>
A;Cross-references: EMBL:X58709; GB:S40537; NID:g35659
A;Note: this translation is not annotated in GenBank entry HSFCOEB1, release 111.0
R;Nunez, A.M.; Kohno, K.; Martin, G.R.; Yamada, Y.
Gene 44, 11-16, 1986
A;Title: Promoter region of the human pro-alpha-1-(II)-collagen gene.
A;Reference number: A24828; MUID: 87031574; PMID: 3021582
A;Accession: A24828
A;Molecule type: DNA
A;Residues: 1-8 'T', 10-28 <NUN>
A;Cross-references: GB:M25698; NID:g180872; PIDN:AAAS2051.1; PID:g553237
R;Baldwin, C.T.; Reginato, A.M.; Smith, C.; Jimenez, S.A.; Prockop, D.J.
Biochem. J. 262, 521-528, 1989
A;Title: Structure of cDNA clones coding for human type II procollagen. The alpha-1(II)
A;Reference number: S06496; MUID: 90026318; PMID: 2803288
A;Accession: S06496
A;Molecule type: mRNA
A;Residues: 7-28 'R', 99-157, 'P', 159-440, 'G', 442-456, 'E', 458-640, 'A', 642-831, 'PA', 834, 'P'

A;Cross-references: EMBL:X16711; NID:g30040; PIDN:CAA34683.1; PID:g30041
A;Note: alternative splice form 1
R;Ryan, M.C.; Sandell, L.J.
J. Biol. Chem. 265, 10334-10339, 1990
A;Title: Differential expression of a cysteine-rich domain in the amino-terminal propept
A;Reference number: A35428; MUID:90285153; PMID:2355003
A;Accession: A35428
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 27-81, 'L', 83-103 <RYA2>
A;Notes: alternative splice form 2; splicing appears to be under developmental regulation
R;Su, M.W.; Benson-Chanda, V.; Vissing, H.; Ramirez, F.
Genomics 4, 438-441, 1989
A;Title: Organization of the exons coding for Pro alpha-1(II) collagen N-propeptide confi
A;Reference number: A30147; MUID:89233138; PMID:2714801
A;Accession: A30147
A;Molecule type: DNA
A;Residues: 104-157, 'P', 159-236 <SUM>
A;Cross-references: GB:J03065; GB:M23660; GB:M25655; GB:M25656; GB:M25730; GB:M32168; GB:G
B;Ala-Tokko, L.; Baldwin, C.T.; Moskowitz, R.W.; Prockop, D.J.
Proc. Natl. Acad. Sci. U.S.A. 87, 6565-6568, 1990
A;Title: Single base mutation in the type II procollagen gene (COL2A1) as a cause of pri
A;Reference number: A94227; MUID:90370826; PMID:1975693
A;Accession: A33116
A;Molecule type: DNA
A;Residues: 171-172, 'C', 174-175 <ALA>
A;Note: mutant sequence from a family with primary generalized osteoarthritis
R;Diab, M.; Wu, J.J.; Eyre, D.R.
Biochem. J. 314, 327-332, 1996
A;Title: Collagen type IX from human cartilage: a structural profile of intermolecular c
A;Reference number: S64673; MUID:96195147; PMID:8660302
A;Accession: S64674
A;Molecule type: protein
A;Residues: 158-189, 'X', 191-195; 1224-1230, 'X', 1232-1236 <DIA>
R;Franc, S.; Marzin, E.; Boutillon, M.M.; Lafont, R.; Lechene de la Porte, P.; Herbage,
Eur. J. Biochem. 234, 125-131, 1995
A;Title: Immunohistochemical and biochemical analyses of 20000-25000-year-old fossil car
A;Reference number: S63514; MUID:96096730; PMID:8529631
A;Accession: S63514
A;Molecule type: protein
A;Residues: 243-261; 575-590; 756-763, 'X', 765-779 <FRA>
R;Tiller, G.E.; Weis, M.A.; Polunbo, P.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre,
Am. J. Hum. Genet. 56, 388-395, 1995
A;Title: An RNA-splicing mutation (G-51VS20) in the type II collagen gene (COL2A1) in a
A;Reference number: I38867; MUID:95150028; PMID:7847372
A;Accession: I38867
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 440, 'G', 442-455, 'E', 458-480, 'P', 482-509 <TIL1>
A;Cross-references: EMBL:U15195; NID:9557053; PIDN:AAB60370.1; PID:g557054
R;Ramirez, F.
submitted to the EMBL Data Library, December 1988
A;Reference number: S04892
A;Accession: S04892
A;Molecule type: mRNA
A;Residues: 501-676, 'A', 678-783, 'A', 785-831, 'EA', 834, 'F', 836-1214 <RAM>
A;Cross-references: EMBL:X13783; NID:g30037; PIDN:CAA32030.1; PID:g930050
R;Vikkula, M.; Peltonen, L.
FEBS Lett. 250, 171-174, 1989
A;Title: Structural analyses of the polymorphic area in type II collagen gene.
A;Reference number: S05000; MUID:89325561; PMID:2753125
A;Accession: S05000
A;Molecule type: DNA
A;Residues: 630-640, 'A', 642-785 <VIK2>
A;Cross-references: EMBL:X16158; NID:ig29951; PIDN:CAA34278.1; PID:g1335018; PIDN:CAA3427
PIDN:CAA34283.1; PID:g1335023; PIDN:CAA34284.1; PID:g1335024
R;Bogaert, R.; Tiller, G.E.; Weis, M.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre, D
J. Biol. Chem. 267, 22522-22526, 1992
A;Title: An amino acid substitution (Gly853--Glu) in the collagen alpha 1(II) chain pro
A;Reference number: A44309; MUID:93054548; PMID:1429602
A;Accession: A44309
A;Status: nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: DNA; mRNA

A:Residues: 752-831, 'PA', 834, 'F', 836-1005, 'K', 1007-1036, 'Q', 1038-1052, 'E', 1054-1068, 'T', 'A'; Cross-references: GB:L00977; NID:G180812; PIDN:AA23914.1; PID:G258774
 A:Note: sequence extracted from NCBI backbone (NCBIP:117273); parts of this sequence were not included in GenBank
 A:Note: this translation is not annotated and this publication is not cited in GenBank
 A:Note: mutant sequence associated with perinatal lethal hypochondrogenesis
 R:Tiller, G.E.; Rimo, D.L.; Murray, L.W.; Cohn, D.H.
 Proc. Natl. Acad. Sci. U.S.A. 87, 3889-3893, 1990
 A:Title: Tandem duplication within a type II collagen gene (COL2A1) exon in an individual
 A:Reference number: S16502; MUID:90251662; PMID:2339128
 A:Accession: S16502
 A:Molecule type: DNA
 A:Residues: 1164-1184, 'GPSKDGANGIPGPI', 1185-1199, 'TIL2'
 A:Cross-references: ENBL:M37126; NID:G180808; PIDN:AA52037.1; PID:G180809
 A:Note: mutant sequence from a patient with spondyloepiphyseal dysplasia
 R:Cheah, K.S.E.; Stoker, N.G.; Griffin, J.R.; Grosfeld, F.G.; Solomon, E.
 Proc. Natl. Acad. Sci. U.S.A. 82, 2555-2559, 1985
 A:Title: Identification and characterization of the human type II collagen gene (COL2A1)
 A:Reference number: A02858; MUID:85190534; PMID:3857598
 A:Accession: A02858
 A:Molecule type: DNA
 A:Residues: 1032-1056, 'N', 1058-1068, 'T', 1070-1487, 'CHE'
 A:Cross-references: GB:J00116; NID:G180395; PIDN:AA51997.1; PID:G180396
 R:Elima, K.; Vuorio, T.; Vuorio, E.
 Nucleic Acids Res. 15, 9499-9504, 1987
 A:Title: Determination of the single polyadenylation site of the human pro-alpha-1(II) chain
 A:Reference number: A27280; MUID:8806771; PMID:2825137
 A:Accession: A27280
 A:Molecule type: DNA; mRNA
 A:Residues: 1175-1487, 'ELI'
 A:Cross-references: ENBL:X06268; NID:G30096; PIDN:CAA29604.1; PID:G30097
 A:Experimental source: fetal epiphyseal cartilage
 R:van der Rest, M.; Rosenberg, L.C.; Olsen, B.R.; Poole, A.R.
 Biochem. J. 237, 923-925, 1986
 A:Title: Chondrocalcin is identical with the C-propeptide of type II procollagen.
 A:Reference number: A57033; MUID:87099927; PMID:3800925
 A:Accession: A57033
 A:Molecule type: protein
 A:Residues: 'X', 1244-1246, 'N', 1248, 'X', 1250-1265; 1295-1305; 1395-1408, 'VAN'
 A:Note: chondrocalcin identified as released collagen I(II) chain carboxyl-terminal propeptide
 R:Strom, C.M.; Upholt, W.B.
 Nucleic Acids Res. 12, 1025-1038, 1984
 A:Title: Isolation and characterization of genomic clones corresponding to the human type II collagen gene
 A:Reference number: A21733; MUID:84118798; PMID:6320112
 A:Accession: A21733
 A:Molecule type: DNA
 A:Residues: 1245-1295, 'STR1'
 A:Cross-references: ENBL:X00339; ENBL:X00298; NID:G394699; PIDN:CAA25092.1; PID:G4378975
 A:Accession: B21733
 A:Molecule type: DNA
 A:Residues: 894-909, 'PE', 'STR2'
 A:Cross-references: GB:X01785; NID:G30035; PIDN:CAA25082.1; PID:G1335032
 R:Nunez, A.M.; Francomano, C.; Young, M.F.; Martin, G.R.; Yamada, Y.
 Biochemistry 24, 6343-6348, 1985
 A:Title: Isolation and partial characterization of genomic clones coding for a human procollagen gene
 A:Reference number: A24561; MUID:86104139; PMID:3002437
 A:Accession: A24561
 A:Molecule type: DNA
 A:Residues: 1296-1358, 'NUN2'
 A:Cross-references: GB:M12048; NID:G180017
 A:Note: this translation is not annotated in GenBank entry HUMCCT2A, release 111.0
 R:Sangiorji, F.O.; Benson-Chanda, V.; de Wet, W.J.; Sobel, M.E.; Tsipouras, P.; Ramirez, N.
 Nucleic Acids Res. 13, 2207-2225, 1985
 A:Title: Isolation and partial characterization of the entire human pro alpha 1(II) collagen gene
 A:Reference number: I37249; MUID:85215609; PMID:2987845
 A:Accession: S59491
 A:Molecule type: DNA
 A:Residues: 7'28', 'R', 99-114; 541-578; 786-802; 1055-1056, 'N', 1058-1068, 'T', 1070-1109; 1200-1201
 A:Accession: 184453
 A:Status: translated from GB/ENBL/DBU
 A:Molecule type: DNA
 A:Residues: 7-28, 'SAN2'

A:Cross-references: GB:M23759; NID:G180845; ENBL:X03320; GB:M24938; NID:G30104
 A:Note: the GenBank PID is based on an incorrect reading frame
 A:Accession: I37250
 A:Status: translated from GB/ENBL/DBU
 A:Molecule type: DNA
 A:Residues: 541-560, 'SAN3'
 A:Cross-references: ENBL:X02378; GB:M23870; NID:G30107; PIDN:CAA26227.1; PID:G929621
 A:Accession: I37251

Query Match 72.8%; Score 862; DB 1; Length 1487;
 Best Local Similarity 74.0%; Pred. No. 5, 1e-41;
 Matches 154; Conservative 11; Mismatches 44; Indels 0; Gaps 0;
 QY 1 GPPGPGTGLPGLPGRGGPGRGPGADGVAGPKGPGAGRGQGGPAGSGPFGQLPGPA 60
 Db 492 GARGPGGVGPIGPPGPGAPGNRFPQDGLAGKAGPGRGPGSLAGKANGDPGRP 551
 QY 61 GEAGLPKAGLGTSGSPGPGKTPGPGAGDGRPPGPGGPGGARGQAGVGMFFGPKGAA 120
 Db 552 GEPGLPGARGLTGRPDAGPQGVGPGAPGDPGPPGPGGARGQGVGMFFGPKGAN 611
 QY 121 GEPGKAGRGVPPGAVGPGKDGACAGAGQPPGPGAGRGQGGPAGSGPFGQLPGPA 180
 Db 612 GEPGKAGKGLPGAPGLRGLPGKGTGAGGPGGPGAGRGQGGPAGSGPFGQLPGP 671
 QY 181 GPPGAGKPGGQGVGPDGLGAPGSPGAG 208
 Db 672 GPPGKPGQGVGPGAGPGLVPRG 699

RESULT 6

T45467
 collagen alpha 1(II) chain precursor [imported] - horse
 N:Alternate names: type II collagen
 C:Species: Equus caballus (domestic horse)
 C:Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 04-Mar-2000
 C:Accession: T45467
 R:Richardson, D.W.; Dodge, G.R.
 submitted to the ENBL Data Library, June 1996
 A:Description: Cloning of equine type II collagen and modulation of its expression in eq
 A:Reference number: Z22977
 A:Accession: T45467
 A:Status: preliminary; translated from GB/ENBL/DBU
 A:Molecule type: mRNA
 A:Residues: 1-1418, 'RIC'
 A:Cross-references: ENBL:U62528; PIDN:AA805773.1
 C:Superfamily: collagen alpha 1(II) chain; fibrillar collagen carboxyl-terminal homology;

Query Match 72.6%; Score 860; DB 2; Length 1418;
 Best Local Similarity 74.0%; Pred. No. 6, 4e-41;
 Matches 154; Conservative 10; Mismatches 44; Indels 0; Gaps 0;
 QY 1 GPPGPGTGLPGLPGRGGPGRGPGADGVAGPKGPGAGRGQGGPAGSGPFGAGRP 60
 Db 423 GARGPGGVGPIGPPGPGAPGNRFPQDGLAGKAGPGRGPGSLAGKANGDPGRP 482
 QY 61 GEAGLPKAGLGTSGSPGPGKTPGPGAGDGRPPGPGGARGQAGVGMFFGPKGAA 120
 Db 483 GEPGLPGARGLTGRPDAGPQGVGPGAPGDPGPPGPGGARGQGVGMFFGPKGAN 542
 QY 121 GEPGKAGRGVPPGAVGPGKDGACAGAGQPPGPGAGRGQGGPAGSGPFGQLPGPA 180
 Db 543 GEPGKAGKGLPGAPGLRGLPGKGTGAGGPGGPGAGRGQGGPAGSGPFGQLPGP 602
 QY 181 GPPGAGKPGGQGVGPDGLGAPGSPGAG 208
 Db 603 GPPGKPGQGVGPGAGPGLVPRG 630

RESULT 7

B40333
 collagen alpha 1(II) chain precursor - African clawed frog
 C:Species: Xenopus laevis (African clawed frog)

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C/Accession: B40333
 R/Su, M.W.; Suzuki, H.R.; Bleker, J.J.; Solursh, M.; Ramirez, F.
 J. Cell Biol. 115, 565-575, 1991
 A/Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis em
 F:37-96/Domain: von Willebrand factor type C repeat homology <VMC>
 A/Reference number: A40333; MUID:92011898; PMID:1918153
 A/Accession: B40333
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-1486 <SUA>
 A/Cross-references: GB:M63595
 C/Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 C/Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
 F:1258-1486/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
 Query Match 71.8%; Score 850; DB 1; Length 1486;
 Best Local Similarity 73.1%; Pred. No. 2.4e-40;
 Matches 152; Conservative 11; Mismatches 45; Indels 0; Gaps 0;
 QY 1 GPPGEPGPTGLPGRGSGRGGFPAGDGVAGPKGAPGERSGPGAGPKSGFGEAGRP 60
 DB 494 GARGEPGAAGPLGPPGERGAPGNRFFPGDGLAGPKGAPGERSGPGAGPKSGFGEAGRP 553
 QY 61 GEAGLPKAGLGTSGPSGPDGKTPGPPAGODGCRPGPPPGARGQAGVWGPFGPKGAA 120
 DB 554 GEPGLPARGLTGRPDAGPQGVKVPAGSGEDGRPPFPQAGRGQPGVWGPFGPKGAN 613
 QY 121 GEPKAGRGVPPGPPGAVGAPKADGEAGAGQPPGAPGAGRGQPGAGPGFGLPGPA 180
 DB 614 GEPKAGKGLVGLAPGLRGLPKDGTGSGQPNPGAPGAGRGQPGAGPGFGLPGPP 673
 QY 181 GPPGKAGRGVPPGPDGLGAPGSGPAG 208
 DB 674 GSPGEGKPGDQGVFGAGAPGLVGRG 701
 RESULT 8
 A40333
 collagen alpha 1'(II) chain precursor - African clawed frog
 C/Species: Xenopus laevis (African clawed frog)
 C/Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 16-Jul-1999
 C/Accession: A40333
 R/Su, M.W.; Suzuki, H.R.; Bleker, J.J.; Solursh, M.; Ramirez, F.
 J. Cell Biol. 115, 565-575, 1991
 A/Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis em
 F:37-96/Domain: von Willebrand factor type C repeat homology <VMC>
 A/Reference number: A40333; MUID:92011898; PMID:1918153
 A/Accession: A40333
 A/Status: nucleic acid sequence not shown
 A/Molecule type: mRNA
 A/Residues: 1-1492 <SUA>
 A/Cross-references: GB:M63596
 A/Note: this sequence is presented as substitutions relative to another sequence in a fi
 es they replace; the appropriate interpretation of the sequence figure was reconstructed
 C/Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 C/Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
 F:1263-1492/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
 Query Match 71.5%; Score 847; DB 2; Length 1492;
 Best Local Similarity 73.6%; Pred. No. 3.5e-40;
 Matches 153; Conservative 9; Mismatches 46; Indels 0; Gaps 0;
 QY 1 GPPGEPGPTGLPGRGSGRGGFPAGDGVAGPKGAPGERSGPGAGPKSGFGEAGRP 60
 DB 497 GARGEPGAAGPLGPPGERGAPGNRFFPGDGLAGPKGAPGERSGPGAGPKSGFGEAGRP 556
 QY 61 GEAGLPKAGLGTSGPSGPDGKTPGPPAGODGCRPGPPPGARGQAGVWGPFGPKGAA 120
 DB 557 GEPGLPARGLTGRPDAGPQGVKVPAGSGAEDGRPPFPQAGRGQPGVWGPFGPKGAN 616
 QY 121 GEPKAGRGVPPGPPGAVGAPKADGEAGAGQPPGAPGAGRGQPGAGSGFGLPGPA 180

617 GPPGKAGSKGLGAPGLRLGLPKDGTGAQGNPGAPGAGRGQPGPSGFGQLPGPP 676
 QY 181 GPPGKAGSKGLGAPGLRLGLPKDGTGAQGNPGAPGAGRGQPGPSGFGQLPGPP 676
 DB 677 GSPGEGKPGDQGVFGAGAGLVGRG 704
 RESULT 9
 A41182
 collagen alpha 1(II) chain precursor - mouse
 C/Species: Mus musculus (house mouse)
 C/Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 13-Aug-1999
 C/Accession: A41182; A44885
 R/Metsaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.
 J. Biol. Chem. 266, 16862-16869, 1991
 A/Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and
 A/Reference number: A41182; MUID:91358489; PMID:1885613
 A/Accession: A41182
 A/Status: preliminary; not compared with conceptual translation
 A/Molecule type: DNA
 A/Residues: 1-1419 <MET>
 A/Cross-references: GB:M65161
 R/Chen, K.S.; Lau, E.T.; Au, P.K.; Tam, P.P.
 Development 111, 945-953, 1991
 A/Title: Expression of the mouse alpha 1(II) collagen gene is not restricted to cartilag
 A/Reference number: A44885; MUID:91347939; PMID:1879363
 A/Accession: A44885
 A/Molecule type: DNA
 A/Residues: 1-28 <CHE>
 A/Cross-references: GB:S63190; NID:9234368; PIDN:AAB19627.1; FID:9234369
 A/Note: sequence extracted from NCBI backbone (NCBI:63190, NCBI:63192)
 C/Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 C/Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trimer;
 F:1191-1419/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
 Query Match 70.9%; Score 839; DB 2; Length 1419;
 Best Local Similarity 72.1%; Pred. No. 9.3e-40;
 Matches 150; Conservative 12; Mismatches 46; Indels 0; Gaps 0;
 QY 1 GPPGEPGPTGLPGRGSGRGGFPAGDGVAGPKGAPGERSGPGAGPKSGFGEAGRP 60
 DB 424 GARGEPGAGPLGPPGERGAPGNRFFPGDGLAGPKGAPGERSGPGAGPKSGFGEAGRP 483
 QY 61 GEAGLPKAGLGTSGPSGPDGKTPGPPAGODGCRPGPPPGARGQAGVWGPFGPKGAA 120
 DB 484 GEPGLPARGLTGRPDAGPQGVKVPAGSGEDGRPPPGQARGQPGVWGPFGPKGAN 543
 QY 121 GSPGKAGRGVPPGPPGAVGAPKADGEAGAGQPPGAPGAGRGQPGAGSGFGLPGPA 180
 DB 544 GEPGKAGRGVPPGPPGAVGAPGLRLGLPKDGTGAAGPPGSGPAGRGQPGAGPSGFGQLPGPP 603
 QY 181 GPPGKAGSKGLGAPGLRLGLPKDGTGAQGNPGAPGAGRGQPGPSGFGQLPGPP 676
 DB 604 GPPGEGKPGDQGVFGAGAGLVGRG 631
 RESULT 10
 B41182
 collagen alpha 1(II) chain precursor (long splice form) - mouse
 C/Species: Mus musculus (house mouse)
 C/Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 16-Jul-1999
 C/Accession: B41182
 R/Metsaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.
 J. Biol. Chem. 266, 16862-16869, 1991
 A/Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and
 A/Reference number: A41182; MUID:91358489; PMID:1885613
 A/Accession: B41182
 A/Status: preliminary; not compared with conceptual translation
 A/Molecule type: DNA
 A/Residues: 1-1487 <MET>
 A/Cross-references: GB:M65161
 C/Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 C/Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trimer;

J. Biol. Chem. 252, 639-643, 1977
A:Title: The covalent structure of cartilage collagen. Evidence for sequence heterogeneity
A:Reference number: A92210; MUID:7703864; PMID:833147
A:Accession: A92210
A:Molecule type: protein
A:Residues: 139-178, 'Z', 180-184, 'PA', 187-190, 'AS', 193-194, 'T', 196-198 <BU2>
A:Cross-references: EMBL:U07973; NID:9520454; PIDN:AAA83407.1; PID:G537432
A:Status: preliminary; translated from GE/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-886 <NAH>
A:Cross-references: EMBL:U07973; NID:9520454; PIDN:AAA83407.1; PID:G537432
C:Genetics:
A:Gene: COL1A1
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
F:30-90/Domain: von Willebrand factor type C repeat homology <WVC>
Query Match 67.1%; Score 794.5; DB 2; Length 886;
Best Local Similarity 65.9%; Pred. No. 1.9e-37;
Matches 143; Conservative 13; Mismatches 52; Indels 9; Gaps 1;
QY 1 GPP-----GEPGPTGLPFPGERGSGRSGFPGADGVAGPKGAPAGERGSPGAPGPK 51
DB 458 GPPGEGKRGANGEPQGVPTGERSGPGRLGNSGLPGKGPAGERGSPGPPGSP 517
QY 52 GSPGAGRPGEAGLPGAKGLTSGSPGPDGKTGTPGAGQDGRPGPPGARGQAGVM 111
DB 518 GPAGRGQDGGFGLPGMGLPGIPGSPGSDGKPGPGNQGEPGRSGPPGAPGPPGQV 577
QY 112 GPPGPKGAAGEPGKAGRGVPGPGVAGVAGKDGAGAGQGGPPGAPAGERGEOGAGSP 171
DB 578 GPPGPKGNEGAFKNGERGPGGPGTGPAGKNGDVGLPGPPGAPAGDRGEPGPGSP 637
QY 172 GFQGLPGPAGPFGHAGKPGEQVPGDGLGAPGSPGAG 208
DB 638 GLQGLPGPGPAGENGKPGEPKPGDGGPGFPKPG 674
RESULT 14
CGH2V
collagen alpha 2(V) chain precursor - human
C:Species: Homo sapiens (man)
C:Date: 31-Jul-1989 #sequence revision 28-Jul-1995 #text change 31-Dec-2000
C:Accession: A31427; A54555; S43643; A25874; I55239; I59025; A25374; A30017
R:Woodbury, D.; Benson-Chanda, V.; Ramirez, F.
J. Biol. Chem. 264, 2735-2738, 1989
A:Title: Amino-terminal propeptide of human pro-alpha2(V) collagen conforms to the structure of alpha2(V) chain precursor
A:Reference number: A31427; MUID:89123368; PMID:2914927
A:Accession: A31427
A:Molecule type: mRNA
A:Residues: 1-463 <WOO>
A:Cross-references: GB:J04478; NID:gl79697; PIDN:AAA51859.1; PID:gl79698
A:Experimental source: Placenta
R:Greenspan, D.S.; Lee, S.T.; Lee, B.S.; Hoffman, G.G.
Gene Expr. 1, 29-39, 1991
A:Title: Homology between alpha2(V) and alpha1(III) collagen promoters and evidence for
A:Reference number: A54555; MUID:92314691; PMID:1820205
A:Accession: A54555
A:Molecule type: DNA
A:Residues: 1-32 <GRE>
A:Cross-references: GB:M58529; NID:gl80834; PIDN:AAC41699.1; PID:G553235
R:Moradi-Ameili, M.; Rousseau, J.C.; Klemm, J.P.; Champilaud, M.F.; Boutillon, M.M.; Berr
Eur. J. Biochem. 221, 987-995, 1994
A:Title: Diversity in the processing events at the N-terminus of type-V collagen.
A:Reference number: S43642; MUID:94237164; PMID:8181482
A:Accession: S43642
A:Molecule type: protein
A:Residues: 288-291, 'P', 293-294, 'X', 296-297, 606, 'X', 608-617 <MOR>
R:Weil, D.; Bernard, M.; Gargano, S.; Ramirez, F.
Nucleic Acids Res. 15, 181-198, 1987
A:Title: The pro alpha 2(V) collagen gene is evolutionarily related to the major fibrillar
A:Reference number: A25874; MUID:87146331; PMID:3029669
A:Accession: A25874
A:Molecule type: mRNA; DNA
A:Residues: 398-1496 <WEI>
A:Cross-references: GB:X04758; NID:G29588; PIDN:CAA28454.1; PID:gl340175
A:Experimental source: rhabdomyosarcoma cell line
R:Myers, J.C.; Loidl, H.R.; Stolle, C.A.; Seyer, J.M.
J. Biol. Chem. 260, 5533-5541, 1985

J. Biol. Chem. 252, 639-643, 1977
A:Title: The covalent structure of cartilage collagen. Evidence for sequence heterogeneity
A:Reference number: A92210; MUID:7703864; PMID:833147
A:Accession: A92210
A:Molecule type: protein
A:Residues: 139-178, 'Z', 180-184, 'PA', 187-190, 'AS', 193-194, 'T', 196-198 <BU2>
A:Cross-references: EMBL:U07973; NID:9520454; PIDN:AAA83407.1; PID:G537432
A:Status: preliminary; translated from GE/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-886 <NAH>
A:Cross-references: EMBL:U07973; NID:9520454; PIDN:AAA83407.1; PID:G537432
C:Genetics:
A:Gene: COL1A1
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
F:30-90/Domain: von Willebrand factor type C repeat homology <WVC>
Query Match 67.1%; Score 794.5; DB 2; Length 886;
Best Local Similarity 65.9%; Pred. No. 1.9e-37;
Matches 143; Conservative 13; Mismatches 52; Indels 9; Gaps 1;
QY 1 GPP-----GEPGPTGLPFPGERGSGRSGFPGADGVAGPKGAPAGERGSPGAPGPK 51
DB 458 GPPGEGKRGANGEPQGVPTGERSGPGRLGNSGLPGKGPAGERGSPGPPGSP 517
QY 52 GSPGAGRPGEAGLPGAKGLTSGSPGPDGKTGTPGAGQDGRPGPPGARGQAGVM 111
DB 518 GPAGRGQDGGFGLPGMGLPGIPGSPGSDGKPGPGNQGEPGRSGPPGAPGPPGQV 577
QY 112 GPPGPKGAAGEPGKAGRGVPGPGVAGVAGKDGAGAGQGGPPGAPAGERGEOGAGSP 171
DB 578 GPPGPKGNEGAFKNGERGPGGPGTGPAGKNGDVGLPGPPGAPAGDRGEPGPGSP 637
QY 172 GFQGLPGPAGPFGHAGKPGEQVPGDGLGAPGSPGAG 208
DB 638 GLQGLPGPGPAGENGKPGEPKPGDGGPGFPKPG 674
RESULT 14
CGH2V
collagen alpha 2(V) chain precursor - human
C:Species: Homo sapiens (man)
C:Date: 31-Jul-1989 #sequence revision 28-Jul-1995 #text change 31-Dec-2000
C:Accession: A31427; A54555; S43643; A25874; I55239; I59025; A25374; A30017
R:Woodbury, D.; Benson-Chanda, V.; Ramirez, F.
J. Biol. Chem. 264, 2735-2738, 1989
A:Title: Amino-terminal propeptide of human pro-alpha2(V) collagen conforms to the structure of alpha2(V) chain precursor
A:Reference number: A31427; MUID:89123368; PMID:2914927
A:Accession: A31427
A:Molecule type: mRNA
A:Residues: 1-463 <WOO>
A:Cross-references: GB:J04478; NID:gl79697; PIDN:AAA51859.1; PID:gl79698
A:Experimental source: Placenta
R:Greenspan, D.S.; Lee, S.T.; Lee, B.S.; Hoffman, G.G.
Gene Expr. 1, 29-39, 1991
A:Title: Homology between alpha2(V) and alpha1(III) collagen promoters and evidence for
A:Reference number: A54555; MUID:92314691; PMID:1820205
A:Accession: A54555
A:Molecule type: DNA
A:Residues: 1-32 <GRE>
A:Cross-references: GB:M58529; NID:gl80834; PIDN:AAC41699.1; PID:G553235
R:Moradi-Ameili, M.; Rousseau, J.C.; Klemm, J.P.; Champilaud, M.F.; Boutillon, M.M.; Berr
Eur. J. Biochem. 221, 987-995, 1994
A:Title: Diversity in the processing events at the N-terminus of type-V collagen.
A:Reference number: S43642; MUID:94237164; PMID:8181482
A:Accession: S43642
A:Molecule type: protein
A:Residues: 288-291, 'P', 293-294, 'X', 296-297, 606, 'X', 608-617 <MOR>
R:Weil, D.; Bernard, M.; Gargano, S.; Ramirez, F.
Nucleic Acids Res. 15, 181-198, 1987
A:Title: The pro alpha 2(V) collagen gene is evolutionarily related to the major fibrillar
A:Reference number: A25874; MUID:87146331; PMID:3029669
A:Accession: A25874
A:Molecule type: mRNA; DNA
A:Residues: 398-1496 <WEI>
A:Cross-references: GB:X04758; NID:G29588; PIDN:CAA28454.1; PID:gl340175
A:Experimental source: rhabdomyosarcoma cell line
R:Myers, J.C.; Loidl, H.R.; Stolle, C.A.; Seyer, J.M.
J. Biol. Chem. 260, 5533-5541, 1985

R;Janeczko, R.A.; Ramirez, F.
Nucleic Acids Res. 17, 6742, 1989
A>Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collagen.
A;Reference number: S04887; MUID:89386015; PMID:2780304
A;Accession: S04887
A;Molecule type: mRNA
A;Residues: 148-163, 'G', 164-240, 'D', 242-471, 'D', 473-487, 'L', 489, 'S', 491-613, 'Y', 615-634,
A;Cross-references: EMBL:X15333; NID:929545; PIDN:CAA33387.1; PID:G930045
A;Note: the authors' translation of residues 905-932 is inconsistent with the nucleotide
R;Seyer, J.M.; Kang, A.H.
Biochemistry 16, 1158-1164, 1977
A>Title: Covalent structure of collagen: amino acid sequence of cyanogen bromide peptide
A;Reference number: A90399; MUID:77134724; PMID:557335
A;Accession: A90399
A;Molecule type: protein
A;Residues: 'V', 165-225, 229-232, 'P', 234-292, 'D', 294-398 <SEY1>
A;Experimental source: liver
A;Note: sequence corrected by A94562; attachment of 2-O-alpha-D-glucosyl-O-beta-D-galact
R;Seyer, J.M.
submitted to the Atlas, December 1977
A;Reference number: A94562
A;Accession: A94562
A;Molecule type: protein
A;Residues: 'V', 169-225, 229-277, 'A', 279-292, 'D', 294, 'S', 296-398 <SEY2>
A;Experimental source: liver
A;Note: author submitted corrections to A90399
R;Milewicz, D.M.; Witz, A.M.; Smith, A.C.; Manchester, D.K.; Waldstein, G.; Byers, P.H.
Am. J. Hum. Genet. 53, 62-70, 1993
A>Title: Parental somatic and germ-line mosaicism for a multilexon deletion with unusual
ispring.
A;Reference number: I51868; MUID:93304430; PMID:8317500
A;Accession: I51868
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 186-194 <ML>
A;Cross-references: GB:S62925; NID:9386425; PIDN:ADL1937.1; PID:G4261637
R;Chiodo, A.A.; Sillence, D.O.; Cole, W.G.; Bateman, J.F.
Biochem. J. 311, 939-943, 1995
A>Title: Abnormal type III collagen produced by an exon-17-skipping mutation of the COL3
A;Reference number: S59511; MUID:96067614; PMID:7487954
A;Accession: S59511
A;Molecule type: mRNA
A;Residues: 302-423 <CHI>
A;Cross-references: GB:S79877; NID:G1195576; PIDN:AAB35615.1; PID:G1195577
R;Seyer, J.M.; Kang, A.H.
Biochemistry 17, 3404-3411, 1978
A>Title: Covalent structure of collagen: amino acid sequence of five consecutive CNBr pe
A;Reference number: A90414; MUID:79000343; PMID:687591
A;Accession: A90414
A;Molecule type: protein
A;Residues: 399-675, 'N', 677-727 <SEY3>
A;Experimental source: liver
R;Lee, B.; Vitale, E.; Superti-Furga, A.; Steinmann, B.; Ramirez, F.
J. Biol. Chem. 266, 5256-5259, 1991
A>Title: G to T transversion at position +5 of a splice donor site causes skipping of th
A;Reference number: I55349; MUID:91161621; PMID:1672129
A;Accession: I55349
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 537-605 <LEE>
A;Cross-references: GB:M59312; NID:G180815; PIDN:AAA52041.1; PID:G180816
R;Seyer, J.M.; Mainardi, C.; Kang, A.H.
Biochemistry 19, 1583-1589, 1980
A>Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB5 from ty
A;Reference number: A90438; MUID:80198282; PMID:6246925
A;Accession: A90438
A;Molecule type: protein
A;Residues: 728-895, 'A', 897-964 <SEY4>
A;Experimental source: liver
R;Cole, W.G.; Chiodo, A.A.; Lamande, S.R.; Janeczko, R.; Ramirez, F.; Dahl, H.H.M.; Chan
J. Biol. Chem. 265, 17070-17077, 1990
A>Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping an
A;Reference number: A38303; MUID:91009133; PMID:2145268

A;Accession: A38303
A;Molecule type: mRNA
A;Residues: 861-1015 <COL>
A;Cross-references: GB:J05617; GB:M55603; GB:M59227; NID:G180878; PIDN:AAB59383.1; PID:G1
A;Note: a mutant sequence with 942-977 spliced out from a patient with Ehlers-Danlos syn
R;Mankoo, B.S.; Daigleish, R.
Nucleic Acids Res. 16, 2337, 1988
A>Title: Human pro alpha1(III) collagen: cDNA sequence for the 3' end.
A;Reference number: S02119; MUID:88189827; PMID:3357782
A;Accession: S02119
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 950-1018, 'Y', 1020-1183, 'S', 1185-1466 <MAN>
A;Cross-references: EMBL:X06700; NID:G30053; PIDN:CAA29886.1; PID:G30054
R;Seyer, J.M.; Kang, A.H.
Biochemistry 20, 2621-2627, 1981
A>Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB9 from tyr
A;Reference number: A90446; MUID:81208139; PMID:7016180
A;Accession: A90446
A;Molecule type: protein
A;Residues: 965-979, 'A', 981-984, 'PS', 987, 'QN', 990-1096, 'P', 1098-1152, 'AT', 1155, 'S', 1157-
A;Experimental source: liver
R;Joldi, H.R.; Brinker, J.M.; May, M.; Pihlajaniemi, T.; Morrow, S.; Rosenbloom, J.; Mye
Nucleic Acids Res. 12, 9383-9394, 1984
A>Title: Molecular cloning and carboxyl-propeptide analysis of human type III procollagen
A;Reference number: A93551; MUID:85087944; PMID:6096827
A;Accession: A93551
A;Molecule type: mRNA
A;Residues: 1065-1155, 'P', 1157-1466 <LOI>
A;Cross-references: EMBL:X01655; EMBL:X01742; NID:G93584; PIDN:CAA25821.1
R;Miskulin, M.; Daigleish, R.; Klueve-Beckerman, B.; Kennard, S.I.; Tolstoshev, P.; Brant
Biochemistry 25, 1408-1413, 1986
A>Title: Human type III collagen gene expression is coordinately modulated with the type
A;Reference number: I52393; MUID:86187804; PMID:3754462
A;Accession: I52393
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1161-1200 <M18>
A;Cross-references: GB:M13146; NID:G180415; PIDN:AAA52003.1; PID:G180416
R;Emmanuel, B.S.; Cannizzaro, L.A.; Seyer, J.M.; Myers, J.C.
Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985
A>Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm
A;Reference number: I59025; MUID:85216505; PMID:3858826
A;Accession: I79359
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1165-1196 <EMA>
A;Cross-references: GB:M11134; NID:G180417; PIDN:AAA52004.1; PID:G180418
R;Chu, M.L.; Weil, D.; de Wet, W.; Bernard, M.; Sippola, M.; Ramirez, F.
J. Biol. Chem. 260, 4357-4363, 1985
A>Title: Isolation of cDNA and genomic clones encoding human pro-alpha1(III) collagen. P.
A;Reference number: A92516; MUID:85157600; PMID:2579949
A;Accession: A92516
A;Molecule type: DNA
A;Residues: 1176-1240, 'V', 1242-1356, 'P', 1358-1466 <CHU>
A;Cross-references: GB:M10615; GB:M10793; GB:M10795; GB:M10796; GB:M10797; GB
A;Experimental source: liver
A;Note: the authors translated the codon TTC for residue 1057 as Tyr; the codons given fo
ation
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (C
3-hydroxylated. About 15% of the lysines are 5-hydroxylated and some are subsequently O
C;Genetics.
A;Gene: GDB:COL3A1
A;Cross-references: GDB:118729; OMIM:120180
A;Map position: 2q31-2q31
A;Introns: 27/1; 94/3; 111/3; 149/3; 176/3; 554/3; 587/3; 1175/3; 1275/3; 1337/3; 1418/3
A;Note: the list of introns is incomplete; defects in this gene can result in Ehlers-Dan
C;Complex: type III collagen is a homotrimer of monomers initially linked by disulfide br
er of their length, is formed with desmosine cross-links made from lysine and allylsine re
C;Function:
A;Description: structural component of extracellular fibrous polymer that maintains intes
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; Ehlers-Danlos syndrome; extracellular matrix; Glycoprotein; Hyd;

F:1-23/Domain: signal sequence #status predicted <SIG>
 F:24-153/Domain: amino-terminal propeptide #status predicted <PRO>
 F:31-91/Domain: von Willebrand factor type C repeat homology <VMC>
 F:154-1221/Product: collagen alpha 1(III) chain #status predicted <MAT>
 F:154-167/Region: amino-terminal nonhelical telopeptide
 F:168-1196/Region: helical
 F:1091-1093/Region: cell attachment (R-G-D) motif
 F:1197-1221/Region: carboxyl-terminal nonhelical telopeptide
 F:1222-1466/Domain: carboxyl-terminal propeptide #status predicted <CPR>
 F:1238-1466/Domain: fibrillar collagen carboxyl-terminal homology <PCC>
 F:24/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
 F:153-154/Cleavage site: Pro-Gln (procollagen N-endopeptidase) #status predicted
 F:154/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
 F:161,1212/Modified site: allysine (Lys) #status predicted
 F:263,284,860,977,1106/Modified site: 5-hydroxylysine (Lys) #status experimental
 F:263/Binding site: carboxylate (Lys) (covalent) #status experimental
 F:584,1094/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental
 F:948-949/Cleavage site: Gly-Ile (collagenase) #status experimental
 F:1106/Binding site: carboxylate (Lys) (covalent) #status predicted

Query Match	65.5%;	Score 776;	DB 1;	Length 1466;
Best Local Similarity	66.3%;	Pred. No. 2.9e-36;		
Matches 138;	Conservative 12;	Mismatches 58;	Indels 0;	Gaps 0;

QY	1	GPPGEPGPTGLPGPGRGPGSGRGGFGAGGVAGPKGPGAGERSGPGAGPKGSPGEGAGRP	60
DB	468	GSPGEPGANGLPGAAGERGAPGFRGAPGNGIPGKGPAGRGAPGAPGPRGAAGEPGRD	527
QY	61	GEAGLPKAGLGTGSPGPGPDGKTGPPGAGDGRGPGPGPPGARGOAGVGMGPPGPKGAA	120
DB	528	GVPGGPGVGMGPGSGPGSGDKGPPGSGQESGSRPPGPPGPGPGPGPGKND	587
QY	121	GEPKAGERGVPFGVAGVAGKDGAGAGAGGPPGAGPAGERGEGGPGSGFGLPGPA	180
DB	588	GAPGKNGERGPGGPGGPGQPGPKNGETGPGPGFTGPGGDKGDTGPPGPGQGLGLPGTG	647
QY	181	GPPGEGAKPGEGVPGDLGAPGPGSPAG	208
DB	648	GPPGNGKPGEPGPKGAGAGAPGKGK	675

Search completed: September 24, 2004, 11:11:45
 Job time : 9.04914 secs

Query Match 97.0%; Score 1149; DB 3; Length 1341;
Best Local Similarity 93.6%; Pred. No. 2e-72;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEGPTGLPGPPGGRGPGSGRPGADGVAGPKGPAGERGSPGA 48
DB 334 GPPGAGBEGKRGARGEPTGLPGPPGGRGPGSGRPGADGVAGPKGPAGERGSPGA 393
QY 49 GPKSGPGEAGRPGEGAGLPGAKGLTGTSPGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 108
DB 394 GPKSGPGEAGRPGEGAGLPGAKGLTGTSPGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 453
QY 109 GVMGPPGPKGAAGEPKKAGERGVPPGAVGPKDGEAGAGQGGPPGAGPAGERGEGGPA 168
DB 454 GVMGPPGPKGAAGEPKKAGERGVPPGAVGPKDGEAGAGQGGPPGAGPAGERGEGGPA 513
QY 169 GSPFGQGLPGPAGPPGEAGKPGEGQGVPGDLGAPGSPGAG 208
DB 514 GSPFGQGLPGPAGPPGEAGKPGEGQGVPGDLGAPGSPGAG 553

RESULT 2

US-09-500-811-18
; Sequence 18, Application US/09500811
; Patent No. 6323314
; GENERAL INFORMATION:

APPLICANT: Qvist, Per
APPLICANT: Bonde, Martin
TITLE OF INVENTION: A Method for Assaying Collagen Fragments
TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
TITLE OF INVENTION: Disorders Associated with the Metabolism of
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:

ADDRESSEE: Darby & Darby PC
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/500,811
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/187,319
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Gogoris, Adda C
REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 4305/08701
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
TELEX: 236687

INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 1341 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: COLLAGEN ALPHA 1 (I)

US-09-500-811-18

Query Match 97.0%; Score 1149; DB 4; Length 1341;

Query Match 97.0%; Score 1149; DB 4; Length 1341;
Best Local Similarity 93.6%; Pred. No. 2e-72;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEGPTGLPGPPGGRGPGSGRPGADGVAGPKGPAGERGSPGA 48
DB 334 GPPGAGBEGKRGARGEPTGLPGPPGGRGPGSGRPGADGVAGPKGPAGERGSPGA 393
QY 49 GPKSGPGEAGRPGEGAGLPGAKGLTGTSPGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 108
DB 394 GPKSGPGEAGRPGEGAGLPGAKGLTGTSPGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 453
QY 109 GVMGPPGPKGAAGEPKKAGERGVPPGAVGPKDGEAGAGQGGPPGAGPAGERGEGGPA 168
DB 454 GVMGPPGPKGAAGEPKKAGERGVPPGAVGPKDGEAGAGQGGPPGAGPAGERGEGGPA 513
QY 169 GSPFGQGLPGPAGPPGEAGKPGEGQGVPGDLGAPGSPGAG 208
DB 514 GSPFGQGLPGPAGPPGEAGKPGEGQGVPGDLGAPGSPGAG 553

RESULT 3

US-09-570-573-18
; Sequence 18, Application US/09570573
; Patent No. 6342361
; GENERAL INFORMATION:

APPLICANT: Qvist, Per
APPLICANT: Bonde, Martin
TITLE OF INVENTION: A Method for Assaying Collagen Fragments
TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
TITLE OF INVENTION: Disorders Associated with the Metabolism of
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:

ADDRESSEE: Darby & Darby PC
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/570,573
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/187,319
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Gogoris, Adda C
REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 4305/08701
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
TELEX: 236687

INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 1341 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: COLLAGEN ALPHA 1 (I)

US-09-570-573-18

Query Match 97.0%; Score 1149; DB 4; Length 1341;

Best Local Similarity 93.6%; Pred. No. 2e-72; Mismatches 2; Indels 12; Gaps 1;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
QY 1 GPP-----GEPGPTGLPGERGSGRGGFPCGADGVAGPKGPGAGSGPGPA 48
Db. |||||
334 GPPGAGEGKRGARGEPGPTGLPGERGSGRGGFPCGADGVAGPKGPGAGSGPGPA 393
QY 49 GPKGSGEAGRPGEAGLPGAKGLTSGSGSPGDKTTPGPAQDGRPPGPPGARGQA 108
Db. |||||
394 GPKGSGEAGRPGEAGLPGAKGLTSGSGSPGDKTTPGPAQDGRPPGPPGARGQA 453
QY 109 GVMGFFPKGAAGPFGKAGRGVPPGAVGPGAGKDGGAQAQPPGAGPAGERGEGQA 168
Db. |||||
454 GVMGFFPKGAAGPFGKAGRGVPPGAVGPGAGKDGGAQAQPPGAGPAGERGEGQA 513
QY 169 GSPGFGQLPGPAGPGEAGKPGEGVPGDLGAPGSPGAG 208
Db. |||||
514 GSPGFGQLPGPAGPGEAGKPGEGVPGDLGAPGSPGAG 553

RESULT 4
US-09-548-608-18
; Sequence 18, Application US/09548608
; Patent No. 6355442
; GENERAL INFORMATION:
; APPLICANT: Ovist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/548,608
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1341 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (I)
US-09-548-608-18

Query Match 97.0%; Score 1149; DB 4; Length 1341;
Best Local Similarity 93.6%; Pred. No. 2e-72;

Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
QY 1 GPP-----GEPGPTGLPGERGSGRGGFPCGADGVAGPKGPGAGSGPGPA 48
Db. |||||
334 GPPGAGEGKRGARGEPGPTGLPGERGSGRGGFPCGADGVAGPKGPGAGSGPGPA 393
QY 49 GPKGSGEAGRPGEAGLPGAKGLTSGSGSPGDKTTPGPAQDGRPPGPPGARGQA 108
Db. |||||
394 GPKGSGEAGRPGEAGLPGAKGLTSGSGSPGDKTTPGPAQDGRPPGPPGARGQA 453
QY 109 GVMGFFPKGAAGPFGKAGRGVPPGAVGPGAGKDGGAQAQPPGAGPAGERGEGQA 168
Db. |||||
454 GVMGFFPKGAAGPFGKAGRGVPPGAVGPGAGKDGGAQAQPPGAGPAGERGEGQA 513
QY 169 GSPGFGQLPGPAGPGEAGKPGEGVPGDLGAPGSPGAG 208
Db. |||||
514 GSPGFGQLPGPAGPGEAGKPGEGVPGDLGAPGSPGAG 553

RESULT 5
US-09-585-887-9
; Sequence 9, Application US/09585887
; Patent No. 6413742
; GENERAL INFORMATION:
; APPLICANT: Olsen, David R
; APPLICANT: Chang, Robert
; APPLICANT: McMullin, Hugh
; APPLICANT: Hitzeman, Ronald A.
; APPLICANT: Chisholm, George
; TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND
; TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT
; TITLE OF INVENTION: CELLS
; FILE REFERENCE: 225002030400
; CURRENT APPLICATION NUMBER: US/09/585,887
; CURRENT FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/289,578
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: 60/084,828
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1461
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-585-887-9

Query Match 97.0%; Score 1149; DB 4; Length 1461;
Best Local Similarity 93.6%; Pred. No. 2.2e-72;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPGERGSGRGGFPCGADGVAGPKGPGAGSGPGPA 48
Db. |||||
455 GPPGAGEGKRGARGEPGPTGLPGERGSGRGGFPCGADGVAGPKGPGAGSGPGPA 514
QY 49 GPKGSGEAGRPGEAGLPGAKGLTSGSGSPGDKTTPGPAQDGRPPGPPGARGQA 108
Db. |||||
515 GPKGSGEAGRPGEAGLPGAKGLTSGSGSPGDKTTPGPAQDGRPPGPPGARGQA 574
QY 109 GVMGFFPKGAAGPFGKAGRGVPPGAVGPGAGKDGGAQAQPPGAGPAGERGEGQA 168
Db. |||||
575 GVMGFFPKGAAGPFGKAGRGVPPGAVGPGAGKDGGAQAQPPGAGPAGERGEGQA 634
QY 169 GSPGFGQLPGPAGPGEAGKPGEGVPGDLGAPGSPGAG 208
Db. |||||
635 GSPGFGQLPGPAGPGEAGKPGEGVPGDLGAPGSPGAG 674

RESULT 6
US-09-289-578-9
; Sequence 9, Application US/09289578
; Patent No. 6428978
; GENERAL INFORMATION:

APPLICANT: Olsen, David R
APPLICANT: Chang, Robert
APPLICANT: McMullin, Hugh
APPLICANT: Hitzeman, Ronald A.
APPLICANT: Chisholm, George
TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND
TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT
FILE REFERENCE: 225002030400
CURRENT FILING DATE: 1999-04-10
PRIOR FILING DATE: 1999-04-10
PRIOR FILING DATE: 1998-05-08
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 9
LENGTH: 1461
TYPE: PRT
ORGANISM: Homo sapiens
US-09-289-578-9

Query Match 97.0%; Score 1149; DB 4; Length 1461;
Best Local Similarity 93.6%; Pred. No. 2.2e-72;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GFGPTGLPGPPGGRGPGSRGFPAGDVGAGPKGPPGARGQA 48
DB 455 GPPGAGEGKRGARGEPGPTGLPGPPGGRGPGSRGFPAGDVGAGPKGPPGARGQA 514

QY 49 GPKSGPAGRPAGLPGAGLGTGSPGSPGDKTGPAGDGRPPGPPGARGQA 108
DB 515 GPKSGPAGRPAGLPGAGLGTGSPGSPGDKTGPAGDGRPPGPPGARGQA 574

QY 109 GVMGPPGPKGAAGFPKAGRGVPPGAVGPPGAGKDGAGAAQPPGPPGARGQA 168
DB 575 GVMGPPGPKGAAGFPKAGRGVPPGAVGPPGAGKDGAGAAQPPGPPGARGQA 634

QY 169 GSPGFGQLPGPAGPGEAGKPGEGVPGDLGAPGSPGAG 208
DB 635 GSPGFGQLPGPAGPGEAGKPGEGVPGDLGAPGSPGAG 674

RESULT 7
US-09-331-347C-21
; Sequence 21, Application US/09331347C
; Patent No. 6617431
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics, S.A.
; TITLE OF INVENTION: Recombinant Collagens and Derived Proteins Produced by Plants, Me
; TITLE OF INVENTION: obtaining Such and their Uses
; FILE REFERENCE: 1149-3
; CURRENT APPLICATION NUMBER: US/09/331.347C
; CURRENT FILING DATE: 1999-08-17
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-331-347C-21

Query Match 97.0%; Score 1149; DB 4; Length 1464;
Best Local Similarity 93.6%; Pred. No. 2.2e-72;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GFGPTGLPGPPGGRGPGSRGFPAGDVGAGPKGPPGARGQA 48
DB 458 GPPGAGEGKRGARGEPGPTGLPGPPGGRGPGSRGFPAGDVGAGPKGPPGARGQA 517

QY 49 GPKSGPAGRPAGLPGAGLGTGSPGSPGDKTGPAGDGRPPGPPGARGQA 108
DB 518 GPKSGPAGRPAGLPGAGLGTGSPGSPGDKTGPAGDGRPPGPPGARGQA 577

QY 109 GVMGPPGPKGAAGFPKAGRGVPPGAVGPPGAGKDGAGAAQPPGPPGARGQA 168
DB 578 GVMGPPGPKGAAGFPKAGRGVPPGAVGPPGAGKDGAGAAQPPGPPGARGQA 637

QY 169 GSPGFGQLPGPAGPGEAGKPGEGVPGDLGAPGSPGAG 208
DB 638 GSPGFGQLPGPAGPGEAGKPGEGVPGDLGAPGSPGAG 677

RESULT 8
US-08-931-820-1
; Sequence 1, Application US/08931820
; Patent No. 6010863
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Assay for collagen degradation
; NUMBER OF SEQUENCES: 4
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/931.820
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 96202596.1
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1057 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Collagen type I
US-08-931-820-1

Query Match 94.6%; Score 1120; DB 3; Length 1057;
Best Local Similarity 96.2%; Pred. No. 1.7e-70;
Matches 200; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 GPPGEPGPTGLPGPPGGRGPGSRGFPAGDVGAGPKGPPGARGQA 60
DB 309 GPPGAGSAGPFPGLGPGGSGRFPAGDVGAGPKGPPGARGQA 368

QY 61 GEAGLPAGKLTGSPGSPGDKTGPAGDGRPPGPPGARGQA 120
DB 369 GEAGLPAGKLTGSPGSPGDKTGPAGDGRPPGPPGARGQA 428

QY 121 GEPGKAGRGVPPGAVGPPGAGKDGAGAAQPPGPPGARGQA 180
DB 429 GEPGKAGRGVPPGAVGPPGAGKDGAGAAQPPGPPGARGQA 488

QY 181 GPPGAGKPGEGVPGDLGAPGSPGAG 208
DB 489 GPPGAGKPGEGVPGDLGAPGSPGAG 516

RESULT 9
US-09-219-849-48
; Sequence 48, Application US/09219849
; Patent No. 6150981
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOWSWITA, JAN B.

```

; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOOBROEK, ANDREAS
; APPLICANT: WERTEN, MARC W.T.
; APPLICANT: WIND, RICHEL D.
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; TITLE OF INVENTION: PREPARATION THEREOF
; FILE REFERENCE: 2728-2
; CURRENT APPLICATION NUMBER: US/09/219,849
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 48
; LENGTH: 595
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: amino acid sequence
US-09-219-849-48

Query Match          94.3%; Score 1116; DB 3; Length 595;
Best Local Similarity 90.5%; Pred. No. 1.9e-70;
Matches 199; Conservative 4; Mismatches 5; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPPEERGGPSRGGPGADGVAGKXGPKGAGERGSGPPA 48
DB 279 GPPGAGEGKRGARGEPGSGLPPEERGGPSRGGPGADGVAGKXGPKGAGERGSGPPA 338
QY 49 GPKGSPGEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPAGQDGRPPGPPGARGQA 108
DB 339 GPKGSPGEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPAGQDGRPPGPPGARGQA 398
QY 109 GVMGFPFGKGAAGEGPKAGRGVPPGAVGPKAGDGEAGQGGPPGAGPAGERGEGQA 168
DB 399 GVMGFPFGKGTAGBPKAGRGVPPGAVGPKAGDGEAGQGGPPGAGPAGERGEGQA 458
QY 169 GSPGFQGLPGAGPPGEGAGKPGEGQGVPGDLGAPGSPGAG 208
DB 459 GSPGFQGLPGAGPPGEGAGKPGEGQGVPGDLGAPGSPGAG 498

RESULT 11
US-09-219-849-49
; Sequence 49, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOWMSTRA, JAN B.
; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOOBROEK, ANDREAS
; APPLICANT: WERTEN, MARC W.T.
; APPLICANT: WIND, RICHEL D.
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; TITLE OF INVENTION: PREPARATION THEREOF
; FILE REFERENCE: 2728-2
; CURRENT APPLICATION NUMBER: US/09/219,849
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 49
; LENGTH: 822
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: amino acid sequence
US-09-219-849-49

Query Match          94.3%; Score 1116; DB 3; Length 822;
Best Local Similarity 90.5%; Pred. No. 2.6e-70;
Matches 199; Conservative 4; Mismatches 5; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPPEERGGPSRGGPGADGVAGKXGPKGAGERGSGPPA 48
DB 279 GPPGAGEGKRGARGEPGSGLPPEERGGPSRGGPGADGVAGKXGPKGAGERGSGPPA 338
QY 49 GPKGSPGEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPAGQDGRPPGPPGARGQA 108
DB 339 GPKGSPGEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPAGQDGRPPGPPGARGQA 398
QY 109 GVMGFPFGKGAAGEGPKAGRGVPPGAVGPKAGDGEAGQGGPPGAGPAGERGEGQA 168
DB 399 GVMGFPFGKGTAGBPKAGRGVPPGAVGPKAGDGEAGQGGPPGAGPAGERGEGQA 458
QY 169 GSPGFQGLPGAGPPGEGAGKPGEGQGVPGDLGAPGSPGAG 208
DB 459 GSPGFQGLPGAGPPGEGAGKPGEGQGVPGDLGAPGSPGAG 498

RESULT 12

```

US-08-468-996-12
; Sequence 12, Application US/08468996
; Patent No. 6645504
; GENERAL INFORMATION:
; APPLICANT: Weiner, Howard
; APPLICANT: Miller, Ariel
; APPLICANT: Zheng, Zhengi
; APPLICANT: Ahmad, Al-Sabbagh
; TITLE OF INVENTION: BYSTANDER SUPPRESSION OF TYPE 1 DIABETES BY ORAL ADMINISTRATION
; FILE REFERENCE: 1010/16959-US3
; CURRENT APPLICATION NUMBER: US/08/468,996
; CURRENT FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: US 07/843,752
; PRIOR FILING DATE: 1992-02-28
; PRIOR APPLICATION NUMBER: US 07/460,852
; PRIOR FILING DATE: 1990-02-21
; PRIOR APPLICATION NUMBER: US 07/596,936
; PRIOR FILING DATE: 1990-10-15
; PRIOR APPLICATION NUMBER: US 07/065,734
; PRIOR FILING DATE: 1987-06-24
; PRIOR APPLICATION NUMBER: US 07/454,486
; PRIOR FILING DATE: 1989-12-20
; PRIOR APPLICATION NUMBER: US 07/487,732
; PRIOR FILING DATE: 1990-03-02
; PRIOR APPLICATION NUMBER: US 07/551,632
; PRIOR FILING DATE: 1990-07-10
; PRIOR APPLICATION NUMBER: US 07/379,778
; PRIOR FILING DATE: 1989-07-14
; PRIOR APPLICATION NUMBER: US 07/607,826
; PRIOR FILING DATE: 1990-10-31
; PRIOR APPLICATION NUMBER: US 07/595,468
; PRIOR FILING DATE: 1990-10-10
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Bos taurus
US-08-468-996-12

Query Match 76.7%; Score 908; DB 4; Length 492;
Best Local Similarity 76.0%; Pred. No. 4.2e-56;
Matches 165; Conservative 9; Mismatches 25; Indels 18; Gaps 2;

QY 1 GPP-----GEPGPTGLPGPPQERGGPGSRGPPGADGVAGPKGPAGERSGPGA 48
Ddb 280 GPPGPAERKRGKARGEPGPTGLPFPGERGGPPGADGVAGPKGPAGERGAPGA 339
QY 49 GPKGSPGEAGRPCEAGLPGAKGLTUSPGSPGPDGKTGTPPGAGQDGRPGPPGARGQA 108
Ddb 340 GPKGSPGEAGRPCEAGLPGAKGLTUSPGSPGPDGKTGTPPGAGQDGRPGPPGARGQA 399
QY 109 GVMGPPGPKGAGECKACGERGVPPGPGAVGAGKDGAGAGOGPPGAPGACGERGEQGA 168
Ddb 400 GVMGPPGPKGAGECKACGERGVPPGPGAVGAGKDGAGAGOGPPGAPGACGERGEQGA 453
QY 169 GSGPFGQLGPPAGPGEAGKPGEQGVGPDGLGAPGPSG 205
Ddb 454 GADGAPGAPDCKGAGPSGQPCAKGEPDGDAGAKGDAG 490

RESULT 13
US-08-468-996-10
; Sequence 10, Application US/08468996
; Patent No. 6645504
; GENERAL INFORMATION:
; APPLICANT: Weiner, Howard
; APPLICANT: Miller, Ariel
; APPLICANT: Zheng, Zhengi
; APPLICANT: Ahmad, Al-Sabbagh
; TITLE OF INVENTION: BYSTANDER SUPPRESSION OF TYPE 1 DIABETES BY ORAL ADMINISTRATION
; FILE REFERENCE: 1010/16959-US3
; CURRENT APPLICATION NUMBER: US/08/468,996
; CURRENT FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: US 07/843,752
; PRIOR FILING DATE: 1992-02-28
; PRIOR APPLICATION NUMBER: US 07/460,852
; PRIOR FILING DATE: 1990-02-21
; PRIOR APPLICATION NUMBER: US 07/596,936
; PRIOR FILING DATE: 1990-10-15
; PRIOR APPLICATION NUMBER: US 07/065,734
; PRIOR FILING DATE: 1987-06-24
; PRIOR APPLICATION NUMBER: US 07/454,486
; PRIOR FILING DATE: 1989-12-20
; PRIOR APPLICATION NUMBER: US 07/487,732
; PRIOR FILING DATE: 1990-03-02
; PRIOR APPLICATION NUMBER: US 07/551,632
; PRIOR FILING DATE: 1990-07-10
; PRIOR APPLICATION NUMBER: US 07/379,778
; PRIOR FILING DATE: 1989-07-14
; PRIOR APPLICATION NUMBER: US 07/607,826
; PRIOR FILING DATE: 1990-10-31
; PRIOR APPLICATION NUMBER: US 07/595,468
; PRIOR FILING DATE: 1990-10-10
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Bos taurus
US-08-468-996-10

```

;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1060 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Collagen type II
;
US-08-931-820-3
;
Query Match          72.8%; Score 862; DB 3; Length 1060;
Best Local Similarity 74.0%; Pred. No. 1.3e-52;
Matches 154; Conservative 11; Mismatches 43; Indels 0; Gaps 0;
;
QY 1 GPPGEPGPTGLPVPGERGGSGRFFGADGVAGKGPAGERSPPGAPGKSGSFGAGRP 60
DB 311 GARGEPGVGPIPPGERGAFNGRFFGQDGLAGPKGAPGERGFSGLAGPKGANGDFGRP 370
;
QY 61 GEALPGAKGLTGSFGSPGPGKGTGPPGACQDGRPPPPGARGAQGVMGFFGPKGAA 120
DB 371 GEPGLGARGLTGRPDAGPQGVPSGAPGEDGRPPGPGQARGQGVMGFFGPKGAN 430
;
QY 121 GEPKAGERGVPGPAGVAGPAGKDGAGAGQPPGAPGAGRGPGAGSFGFQGLPGPA 180
DB 431 GEPKAGERGVPGPAGVAGPAGKDGAGAGQPPGAPGAGRGPGAGRGQAGPSPGQGLPGPP 490
;
QY 181 GPPGEAGKPGGQGVPGDILGAPSPGAP 208
DB 491 GPPGEGGKPGQGVPGGAGAPGLVPRG 518
;
;
RESULT 15
US-08-963-825-20
; Sequence 20, Application US/08963825
; Patent No. 6110689
;
GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
;
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/963,825
; FILING DATE:
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/187,319
; FILING DATE: 21-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELE: 236687
; INFORMATION FOR SEQ ID NO: 20:

```

```

;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1418 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN -ALPHA 1 (II)
;
US-08-963-825-20
;
Query Match          72.8%; Score 862; DB 3; Length 1418;
Best Local Similarity 74.0%; Pred. No. 1.7e-52;
Matches 154; Conservative 11; Mismatches 43; Indels 0; Gaps 0;
;
QY 1 GPPGEPGPTGLPVPGERGGSGRFFGADGVAGKGPAGERSPPGAPGKSGSFGAGRP 60
DB 423 GARGEPGVGPIPPGERGAFNGRFFGQDGLAGPKGAPGERGFSGLAGPKGANGDFGRP 482
;
QY 61 GEALPGAKGLTGSFGSPGPGKGTGPPGACQDGRPPPPGARGAQGVMGFFGPKGAA 120
DB 483 GEPGLGARGLTGRPDAGPQGVPSGAPGEDGRPPGPGQARGQGVMGFFGPKGAN 542
;
QY 121 GEPKAGERGVPGPAGVAGPAGKDGAGAGQPPGAPGAGRGPGAGSFGFQGLPGPA 180
DB 543 GEPKAGERGVPGPAGVAGPAGKDGAGAGQPPGAPGAGRGPGAGRGQAGPSPGQGLPGPP 602
;
QY 181 GPPGEAGKPGGQGVPGDILGAPSPGAP 208
DB 603 GPPGEGGKPGDQGVPGGAGAPGLVPRG 630
;
;
Search completed: September 24, 2004, 11:09:49
Job time : 10.6042 secs

```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2004, 11:06:55 ; Search time 28.5883 Seconds
(without alignments)
2065.614 Million cell updates/sec

Title: US-10-658-989a-1

Perfect score: 1184

Sequence: 1 GPPGEPPTGLPQPPGRRG.....GEGVFDLCAQPGSPAGG 209

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

- 1: Geneseqp1980s.*
- 2: Geneseqp1980s.*
- 3: Geneseqp2000s.*
- 4: Geneseqp2001s.*
- 5: Geneseqp2002s.*
- 6: Geneseqp2003as.*
- 7: Geneseqp2003bs.*
- 8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1149	97.0	501	4 AAB68057	Aab68057 Amino aci
2	1149	97.0	501	4 AAB02703	Aae02703 Human alp
3	1149	97.0	501	7 ADB84290	Adb84290 Recombina
4	1149	97.0	1057	3 AAY84541	Aay84541 Amino aci
5	1149	97.0	1057	3 AAY84544	Aay84544 A human c
6	1149	97.0	1058	3 AAY84403	Aay84403 Amino aci
7	1149	97.0	1107	2 AAR89472	Aar89472 Collagen/
8	1149	97.0	1107	3 AAY84540	Aay84540 Amino aci
9	1149	97.0	1161	7 ADE87050	Ade87050 Human pan
10	1149	97.0	1169	2 AAR89469	Aar89469 Collagen/
11	1149	97.0	1169	3 AAY84537	Aay84537 Amino aci
12	1149	97.0	1171	2 AAR89470	Aar89470 Collagen/
13	1149	97.0	1171	3 AAY84538	Aay84538 A chimeri
14	1149	97.0	1211	7 ADE87057	Ade87057 Human pan
15	1149	97.0	1226	7 ADE87062	Ade87062 Human pan
16	1149	97.0	1341	2 AAR71701	Aar71701 Collagen
17	1149	97.0	1341	3 AAY96122	Aay96122 Collagen
18	1149	97.0	1341	5 AAE16475	Aae16475 Human col
19	1149	97.0	1341	5 ABB0733	Abb0733 Collagen
20	1149	97.0	1341	5 ABB09625	Abb09625 Amino aci
21	1149	97.0	1388	2 AAR89471	Aar89471 Collagen/
22	1149	97.0	1411	3 AAY56800	Aay56800 Human pre
23	1149	97.0	1461	5 ABG93947	Abg93947 Human pol
24	1149	97.0	1464	2 AAW68485	Aaw68485 Human rec
25	1149	97.0	1464	4 AAB82454	Aab82454 Human pro

26	1149	97.0	1464	4 AAL14136	Aal14136 Human nov
27	1149	97.0	1464	5 ABB90764	Abb90764 Human tum
28	1149	97.0	1464	5 ABB90764	Abb90764 Human pan
29	1149	97.0	1464	6 ABUS4471	Abus4471 Human tum
30	1149	97.0	1464	6 ABR47417	Abx47417 Breast ca
31	1149	97.0	1464	6 ABR92064	Abx92064 Human cer
32	1149	97.0	1464	7 ADD14142	Add14142 Human src
33	1149	97.0	1464	7 ADD45059	Add45059 Human pro
34	1149	97.0	1464	7 ADD45055	Add45055 Human pro
35	1149	97.0	1464	7 ADD45051	Add45051 Human pro
36	1149	97.0	1464	7 ADE87048	Ade87048 Human pan
37	1149	97.0	1464	7 ADE87051	Ade87051 Human pan
38	1144	96.6	1449	4 AAB02535	Aae02535 Porcine a
39	1141	96.4	1388	3 AAY84539	Aay84539 Amino aci
40	1141	96.4	1463	4 AAE02532	Aae02532 Bovine al
41	1128	95.3	1518	4 ABG22879	Abg22879 Novel hum
42	1122	94.8	1453	7 ADD45053	Add45053 Rat Prote
43	1122	94.8	1453	7 ADD45057	Add45057 Rat Prote
44	1122	94.8	1453	7 ADD48341	Add48341 Rat Prote
45	1122	94.8	1453	7 ADD45049	Add45049 Rat Prote

ALIGNMENTS

RESULT 1

AAB68057

ID AAB68057 standard; protein; 501 AA.

AC AAB68057;

DT 09-JUL-2001 (first entry)

DE DE Amino acid sequence of a recombinant human gelatin.

KW Human; gelatin; vaccine; anaphylactic reaction.

XX OS Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 85

FT /note= "this residue is given as unknown as it is

FT illegible in the specification"

XX WO200134801-A2.

PD 17-MAY-2001.

XX 10-NOV-2000; 2000WO-US030843.

XX 12-NOV-1999; 99US-0165114P.

XX 15-MAY-2000; 2000US-0204437P.

XX (FIBR-) FIBROGEN INC.

XX Chang RC, Kivirikko KI, Neff TB, Olsen DR, Polarek JW;

XX WPI; 2001-308784/32.

XX Vaccine formulations (I) comprising recombinant human gelatin, useful for
vaccinating against e.g. mumps, measles, rubella, tetanus, rabies and
cholera, the gelatin is non-immunogenic and confers stability at ambient
temperatures.

XX Claim 11; Page 114-116; 130pp; English.

XX The present sequence represents a human recombinant gelatin polypeptide.
The recombinant gelatin polypeptide is used to produce vaccine
formulations of the invention. The recombinant human gelatin is non-
immunogenic (therefore reducing anaphylactic reactions) and confers
stability at ambient temperatures. The vaccine formulation comprises a
vaccine formulated for the prevention of a disease selected from vaccinia
virus (small pox), polio virus (Salk and Sabin), mumps, measles, rubella,

PR 15-MAY-2000; 2000US-0204437P.
 PR 10-NOV-2000; 2000US-00710249.
 XX (CHAN//) CHANG R C.
 PA (KIVI//) KIVIRIKKO K I.
 PA (NEFF//) NEFF T B.
 PA (OLSE//) OLSEN D R.
 PA (POLA//) POLAREK J W.
 XX Chang RC, Kivirikko KI, Neff TB, Olsen DR, Polarek JW;
 FI WPI; 2003-540775/51.
 XX New vaccine composition comprising a recombinant gelatin and an antigenic
 XX agent, useful for preventing e.g. polio virus, mumps, measles, rubella,
 PT diphtheria, tetanus, chicken pox/shingles, pertussis, cholera, rotavirus
 PT or dengue.
 XX Claim 16; Page 36-38; 63pp; English.
 XX The invention describes a vaccine composition comprising a recombinant
 CC gelatin, and an antigenic agent. The vaccine can be delivered by
 CC injection, through nasal, oral, transdermal or mucosal routes, or through
 CC deep lung delivery. Administration may also be oral, rectal,
 CC transcutaneous, intramedullary, intrathecal, intraventricular,
 CC subcutaneous, intranasal or intracocular injection). The vaccine
 CC composition is formulated for the prevention of vaccinia virus (small
 CC pox), polio virus (Salk and Sabin), mumps, measles, rubella, diphtheria,
 CC tetanus, Varicella-Zoster (chicken pox/shingles), pertussis (whooping
 CC cough), Bacille Calmette-Guerin (BCG, tuberculosis), Haemophilus
 CC influenzae meningitis, rabies, cholera, Japanese encephalitis virus,
 CC Salmonella typhi, Shigella hepatitis A, hepatitis B, adenovirus, yellow
 CC fever, foot and mouth disease, herpes simplex virus, respiratory
 CC syncytial virus, rotavirus, dengue, West Nile virus, Turkey herpes virus
 CC (Marek's disease), influenza, and anthrax. This is the amino acid
 CC sequence of a recombinant gelatin used in the creation of a vaccine
 CC composition of the invention.
 XX Sequence 501 AA;
 SQ
 Query Match 97.0%; Score 1149; DB 7; Length 501;
 Best Local Similarity 93.6%; Pred. No. 3.5e-67;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GPPGPTGLPGRGSGRPGFPGADGVAGPKGPGAGSGSPGA 48
 DB 280 GPPGAGEGKRGARPGFTGLPGPGRGSGRPGFPGADGVAGPKGPGAGSGSPGA 339
 QY 49 GPKGSPGAGRPGAGLPGAKGLTSGSPGPDGKTGPPAGQDGRPGPPPGARGQA 108
 DB 340 GPKGSPGAGRPGAGLPGAKGLTSGSPGPDGKTGPPAGQDGRPGPPPGARGQA 399
 QY 109 GVMGFPKGKAGRPGKAGRGVPGPCAVGPGACKDGEAGAQGPPGAGPAGERGEGPA 168
 DB 400 GVMGFPKGKAGRPGKAGRGVPGPCAVGPGACKDGEAGAQGPPGAGPAGERGEGPA 459
 QY 169 GSPGFQGLPGPAGPGEAGKPGCEQGVPGDLCGAPGSPGAG 208
 DB 460 GSPGFQGLPGPAGPGEAGKPGCEQGVPGDLCGAPGSPGAG 499
 RESULT 4
 ID AAY84541
 XX Amino acid sequence of a human collagen 1 (alpha1) protein.
 XX Extracellular matrix protein; self aggregation; hydroxylated proline;
 AC AAY84541;
 XX
 DT 25-JUL-2000 (first entry)
 XX
 DE Amino acid sequence of a human collagen 1 (alpha1) protein.
 XX
 KW Extracellular matrix protein; self aggregation; hydroxylated proline;

KW trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
 KW collagen; fibrinogen; fibronectin; post translational hydroxylation.
 OS Homo sapiens.
 XX EP992586-A2.
 XX 12-APR-2000.
 XX 07-OCT-1999; 99EP-00119184.
 XX 09-OCT-1998; 98US-00169768.
 XX (USSU) US SURGICAL CORP.
 XX Gruskin EA, Buechter DD, Zhang G, Connolly K;
 XX WPI; 2000-259138/23.
 DR N-PSDB; AAA12502.
 XX Production of extracellular matrix proteins containing 4-trans-
 PT hydroxyproline results in native self aggregating proteins, useful on
 PT medical implants.
 XX Disclosure; Fig 27A-E; 260pp; English.
 XX The specification describes a method for producing an extracellular
 CC matrix protein or its fragment. The extracellular matrix protein is
 CC capable of self aggregating in a cell which does not ordinarily
 CC hydroxylated prolines. The method comprises optimising a nucleic acid
 CC sequence for expression in the cell by substitution of codons preferred
 CC by that cell for naturally occurring codons not preferred by the cell;
 CC incorporating the nucleic acid sequence into the cell; and contacting the
 CC cell with a hypertonic growth medium containing at least one amino acid,
 CC selected from the group consisting of trans-4-hydroxyproline and 3-
 CC hydroxyproline to allow at least one of the amino acids to be assimilated
 CC into the cell and incorporated into the extracellular matrix protein. The
 CC method may be used to make host cells assimilate and incorporate trans-4-
 CC hydroxyproline into proteins. This is especially useful in the
 CC recombinant production of proteins such as collagen, fibrinogen and
 CC fibronectin whose ability to self aggregate and produce functional
 CC proteins depends on the post translational hydroxylation of proline. The
 CC method is also useful in studying the structure and function of
 CC polypeptides which do not normally contain trans-4-hydroxyproline. The
 CC present sequence represents a human collagen 1 (alpha1) protein, which
 CC may be produced using the method of the invention
 XX Sequence 1057 AA;
 SQ
 Query Match 97.0%; Score 1149; DB 3; Length 1057;
 Best Local Similarity 93.6%; Pred. No. 6.4e-67;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GPPGPTGLPGRGSGRPGFPGADGVAGPKGPGAGSGSPGA 48
 DB 297 GPPGAGEGKRGARPGFTGLPGPGRGSGRPGFPGADGVAGPKGPGAGSGSPGA 356
 QY 49 GPKGSPGAGRPGAGLPGAKGLTSGSPGPDGKTGPPAGQDGRPGPPPGARGQA 108
 DB 357 GPKGSPGAGRPGAGLPGAKGLTSGSPGPDGKTGPPAGQDGRPGPPPGARGQA 416
 QY 109 GVMGFPKGKAGRPGKAGRGVPGPCAVGPGACKDGEAGAQGPPGAGPAGERGEGPA 168
 DB 417 GVMGFPKGKAGRPGKAGRGVPGPCAVGPGACKDGEAGAQGPPGAGPAGERGEGPA 476
 QY 169 GSPGFQGLPGPAGPGEAGKPGCEQGVPGDLCGAPGSPGAG 208
 DB 477 GSPGFQGLPGPAGPGEAGKPGCEQGVPGDLCGAPGSPGAG 516
 RESULT 5
 ID AAY84544
 XX Amino acid sequence of a human collagen 1 (alpha1) protein.
 XX Extracellular matrix protein; self aggregation; hydroxylated proline;

XX AC AAY84544;
XX DT 25-JUL-2000 (first entry)
XX DE A human collagen 1 (alpha1) protein helical region.
XX KW Extracellular matrix protein; self aggregation; hydroxylated proline;
XX KW trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
XX KW collagen; fibrinogen; fibronectin; post translational hydroxylation.
XX OS Homo sapiens.
XX PN EP992586-A2.
XX PD 12-APR-2000.
XX PF 07-OCT-1999; 99EP-00119184.
XX PR 09-OCT-1998; 98US-00169768.
XX PA (USSU) US SURGICAL CORP.
XX PI Gruskin EA, Buechter DD, Zhang G, Connolly K;
XX DR WPI: 2000-259138/23.
XX DR N-PSDB; AAA12503.
XX PT Production of extracellular matrix proteins containing 4-trans-
XX PT hydroxyproline results in native self aggregating proteins, useful on
XX PT medical implants.
XX PS Example 10; Fig 39A-E; 260pp; English.
XX CC The specification describes a method for producing an extracellular
XX CC matrix protein or its fragment. The extracellular matrix protein is
XX CC capable of self aggregating in a cell which does not ordinarily
XX CC hydroxylated prolines. The method comprises optimising a nucleic acid
XX CC sequence for expression in the cell by substitution of codons preferred
XX CC by that cell for naturally occurring codons not preferred by the cell;
XX CC incorporating the nucleic acid sequence into the cell; and contacting the
XX CC cell with a hypertonic growth medium containing at least one amino acid,
XX CC selected from the group consisting of trans-4-hydroxyproline and 3-
XX CC hydroxyproline to allow at least one of the amino acids to be assimilated
XX CC into the cell and incorporated into the extracellular matrix protein. The
XX CC method may be used to make host cells assimilate and incorporate trans-4-
XX CC hydroxyproline into proteins. This is especially useful in the
XX CC recombinant production of proteins such as collagen, fibrinogen and
XX CC fibronectin whose ability to self aggregate and produce functional
XX CC proteins depends on the post translational hydroxylation of proline. The
XX CC method is also useful in studying the structure and function of
XX CC polypeptides which do not normally contain trans-4-hydroxyproline. The
XX CC present sequence represents human collagen 1 (alpha1) helical region,
XX CC which may be produced using the method of the invention
XX SQ Sequence 1057 AA;

Query Match 97.0%; Score 1149; DB 3; Length 1057;
Best Local Similarity 93.6%; Pred. No. 6.4e-67;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
QY 1 GPP-----GEPGTLPGPGERGCGSRGPGADGVAGPKGAGERGSGGPA 48
DB 297 GPPGAGEGKRGARGEPGTLPGPGERGCGSRGPGADGVAGPKGAGERGSGGPA 356
QY 49 GPKSGPGEAGRPGEAGLPGAKGLTSGSGSPGDKTGTGPPGAGQDGRPPGPPGARGQA 108
DB 357 GPKSGPGEAGRPGEAGLPGAKGLTSGSGSPGDKTGTGPPGAGQDGRPPGPPGARGQA 416
QY 109 GWMGFPGPKGAAGFPKGAGRGVPGPGAVGPKGDGEAGAQGPPGAGGERGSGGPA 168
DB 417 GWMGFPGPKGAAGFPKGAGRGVPGPGAVGPKGDGEAGAQGPPGAGGERGSGGPA 476

QY 169 GSPGQGLPGPAGPPGEGAKPGEGQGVPGDILGAPGSPGAG 208
DB 477 GSPGQGLPGPAGPPGEGAKPGEGQGVPGDILGAPGSPGAG 516

RESULT 6

AY84403
ID AAY84403 standard; protein; 1058 AA.
XX AC AAY84403;
XX DT 12-JUL-2000 (first entry)
XX DE Amino acid sequence of human type 1 (alpha1) collagen polypeptide.
XX KW Alpha1 collagen; 3,4-dehydro-L-proline; epoxidation; 3,4-epoxyproline;
XX KW collagen; mussel adhesive protein; bioadhesive.
XX OS Homo sapiens.
XX PN WO200014201-A1.
XX PD 16-MAR-2000.
XX PF 07-SEP-1999; 99WO-US020462.
XX PR 09-SEP-1998; 98US-0099652P.
XX PA (USSU) US SURGICAL CORP.
XX PA (PAOL/) PAOLELLA D N.
XX PA (GRUS/) GRUSKIN E A.
XX PA (BUEC/) BUECHTER D D.
XX PI Paolella DN, Gruskin EA, Buechter DD;
XX DR WPI: 2000-271051/23.
XX DR N-PSDB; AAZ99843.
XX PT Incorporating non-natural amino acid into polypeptide, useful e.g. for
XX PT production of bioadhesives, by epoxidation or substitution of
XX PT dehydroproline residues.
XX PS Disclosure; Fig 6; 66pp; English.
XX CC The present sequence represents a human type 1 (alpha1) collagen protein.
XX CC Peptides derived from the protein were used to demonstrate incorporation
XX CC of 3,4-dehydro-L-proline into the peptide, using the method of the
XX CC invention. The specification describes a method for the incorporation of
XX CC non-natural amino acid into a polypeptide. The method comprises reacting
XX CC at least one 3,4-dehydroproline residue in the polypeptide with an
XX CC epoxidation reagent from a polypeptide containing at least one 3,4-
XX CC epoxyproline residue. The method is used for studying the effects of non-
XX CC natural amino acids on structure and function of polypeptides. The method
XX CC is also useful for commercial production of collagen or mussel adhesive
XX CC proteins (which are useful as bioadhesives), and for incorporating a wide
XX CC variety of groups, including therapeutic ligands and biological probes,
XX CC into polypeptides
XX SQ Sequence 1058 AA;

Query Match 97.0%; Score 1149; DB 3; Length 1058;
Best Local Similarity 93.6%; Pred. No. 6.4e-67;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
QY 1 GPP-----GEPGTLPGPGERGCGSRGPGADGVAGPKGAGERGSGGPA 48
DB 298 GPPGAGEGKRGARGEPGTLPGPGERGCGSRGPGADGVAGPKGAGERGSGGPA 357
QY 49 GPKSGPGEAGRPGEAGLPGAKGLTSGSGSPGDKTGTGPPGAGQDGRPPGPPGARGQA 108
DB 358 GPKSGPGEAGRPGEAGLPGAKGLTSGSGSPGDKTGTGPPGAGQDGRPPGPPGARGQA 417
QY 109 GWMGFPGPKGAAGFPKGAGRGVPGPGAVGPKGDGEAGAQGPPGAGGERGSGGPA 168

```

Db 418 GVMGPPGPKGAGEGKAGRGVGPFGAVGPKDGEAGQPPGAGPAGERGEQGP 477
QY 169 GSPFGQLPGPAGGPKGAGEGKQGVPGDLGAPGSPGAG 208
Db 478 GSPFGQLPGPAGGPKGAGEGKQGVPGDLGAPGSPGAG 517

RESULT 7
ID AAR89472
AC AAR89472
XX
XX
XX 01-OCT-1996 (first entry)
DE Collagen/decorin(aa46-93) fusion protein.
KW Transforming growth factor; TGF-beta-1; collagen IA; osteogenesis;
XX bone formation; tissue repair; fusion protein.
XX Synthetic.
OS
XX
XX Key Location/Qualifiers
FH Domain 1..1057
FT /label= Collagen-IA
FT /note= "collagen IA alpha-helical domain"
FT Misc-difference 887 /note= "unidentified amino acid"
FT Misc-difference 890 /note= "unidentified amino acid"
FT Peptide 1058..1059 /label= Linker_peptide
FT Domain 1060..1107 /label= Decorin
FT /note= "amino acids P46 to G93 of mature decorin"
XX
XX CA2151547-A.
XX
XX 11-DEC-1995.
XX
XX 12-JUN-1995; 95CA-02151547.
XX
XX 10-JUN-1994; 94US-00259263.
XX
XX (USSU ) US SURGICAL CORP.
XX
XX Gruskin EA, Espino P;
XX WPI; 1996-140144/15.
XX N-PSDB; AAT16518.
XX
XX Chimaeric DNA encoding protein contg. extracellular matrix protein domain
PT - and cellular regulatory factor domain, partic. useful as osteogenic
PT agents, also related vectors, transformed cells and chimaeric proteins.
XX
XX Disclosure; Fig 8; 59pp; English.
XX
XX A fusion protein (AAR89472) comprises the alpha-helical region of human
XX collagen I(a) linked to amino acids 46-93 of human mature dermatan
XX sulphate proteoglycan (decorin). It can be expressed in Escherichia coli
XX transformants carrying a vector incorporating a chimeric gene (AAT16518)
XX coding for the fusion. The decorin binds to type I collagen and thus
XX affects Eibrl formation. It inhibits the cell attachment-promoting
XX activity of collagen and fibrinogen by binding to such molecules near
XX their cell binding sites. The collagen moiety provides an integral
XX substratum or scaffolding for the decorin. The fusion protein acts to
XX reduce scarring of healing tissue
XX
XX Sequence 1107 AA;
XX
XX Query Match 97.0%; Score 1149; DB 2; Length 1107;
XX Best Local Similarity 93.6%; Pred. No. 6.6e-67;

```

```

Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
QY 1 GPP-----GEPGTGLPGPPGGRGGPFGADGVAGPKGPGAGERSPGPA 48
Db 297 GPPGAGEGKRGAGEGPGTGLPGPPGGRGGPFGADGVAGPKGPGAGERSPGPA 356
QY 49 GPKGSFGEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 108
Db 357 GPKGSFGEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 416
QY 109 GVMGFPKGKAGEGKAGRGVGPFGAVGPKDGEAGAGQPPGAGPAGERGEQGP 168
Db 417 GVMGFPKGKAGEGKAGRGVGPFGAVGPKDGEAGAGQPPGAGPAGERGEQGP 476
QY 169 GSPFGQLPGPAGGPKGAGEGKQGVPGDLGAPGSPGAG 208
Db 477 GSPFGQLPGPAGGPKGAGEGKQGVPGDLGAPGSPGAG 516

RESULT 8
AAR84540
ID AAR84540 standard; protein; 1107 AA.
XX
XX AAR84540;
AC
XX 25-JUL-2000 (first entry)
DE Amino acid sequence of a chimeric collagen 1 (alpha)/decorin protein.
XX
XX Extracellular matrix protein; self aggregation; hydroxylated proline;
KW trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
XX collagen; fibrinogen; fibronectin; post translational hydroxylation;
XX decorin; chimera.
XX
XX Homo sapiens.
OS Unidentified.
OS Chimeric.
XX
XX Key Location/Qualifiers
FH Misc-difference 858 /note= "Gly encoded by GCT"
FT
FT
XX EP992586-A2.
XX
XX 12-APR-2000.
XX
XX 07-OCT-1999; 99EP-00119184.
XX
XX 09-OCT-1998; 98US-00169768.
XX
XX (USSU ) US SURGICAL CORP.
XX
XX Gruskin EA, Buechter DD, Zhang G, Connolly K;
XX WPI; 2000-259138/23.
XX N-PSDB; AAA12500.
XX
XX Production of extracellular matrix proteins containing 4-trans-
PT hydroxyproline results in native self aggregating proteins, useful on
PT medical implants.
XX
XX Claim 24; Fig 18; 260pp; English.
XX
XX The specification describes a method for producing an extracellular
XX matrix protein or its fragment. The extracellular matrix protein is
XX capable of self aggregating in a cell which does not ordinarily
XX hydroxylated prolines. The method comprises optimising a nucleic acid
XX sequence for expression in the cell by substitution of codons preferred
XX by that cell for naturally occurring codons not preferred by the cell;
XX incorporating the nucleic acid sequence into the cell; and contacting the
XX cell with a hypertonic growth medium containing at least one amino acid,
XX selected from the group consisting of trans-4-hydroxyproline and 3-
XX hydroxyproline to allow at least one of the amino acids to be assimilated

```

CC into the cell and incorporated into the extracellular matrix protein. The
 CC method may be used to make host cells assimilate and incorporate trans-4-
 CC hydroxyproline into proteins. This is especially useful in the
 CC recombinant production of proteins such as collagen, fibrinogen and
 CC fibronectin whose ability to self aggregate and produce functional
 CC proteins depends on the post translational hydroxylation of proline. The
 CC method is also useful in studying the structure and function of
 CC polypeptides which do not normally contain trans-4-hydroxyproline. The
 CC present sequence represents a chimeric collagen 1 (alpha1)/decorin
 CC protein, which may be produced using the method of the invention
 XX
 SQ Sequence 1107 AA;

Query Match 97.0%; Score 1149; DB 3; Length 1107;
 Best Local Similarity 93.6%; Pred. No. 6.6e-67;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GPPGPTGLPGERGSGRPGADGVAGPKGAGERGSPGPA 48
 DB 297 GPPGAGEGKRGARPEGPTGLPGERGSGRPGADGVAGPKGAGERGSPGPA 356
 QY 49 GPKSGPGEAGRPGEAGLPGAKLTGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 108
 DB 357 GPKSGPGEAGRPGEAGLPGAKLTGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 416
 QY 109 GVMGFPKGKGAAGSPKAGRGVPPGCVGPKAGDGEAGAGQPPGAGPAGERGEGGPA 168
 DB 417 GVMGFPKGKGAAGSPKAGRGVPPGCVGPKAGDGEAGAGQPPGAGPAGERGEGGPA 476
 QY 169 GSPGFQGLPGPAGPPGEAGKFGQVPGDLGAPGSPGAG 208
 DB 477 GSPGFQGLPGPAGPPGEAGKFGQVPGDLGAPGSPGAG 516

RESULT 9
 ADE87050
 ID ADE87050 standard; protein; 1161 AA.
 AC ADE87050;
 XX
 XX 29-JAN-2004 (first entry)
 DE Human pancreatic cell protein sequence SegID510.
 XX
 XX neoplastic pancreatic cell; pancreatic cancer;
 KW cancer death; cytostatic; vaccine; gene therapy;
 KW non-cancerous pancreas disease; human.
 XX
 OS Homo sapiens.
 XX
 XX WO2003060145-A2.
 PN
 XX 24-JUL-2003.
 PD
 XX 19-DEC-2002; 2002WO-US040655.
 PF
 XX 21-DEC-2001; 2001US-0342768P.
 PR
 XX (DIAD-) DIADEXUS INC.
 PA
 XX Sun Y, Liu C;
 PI
 XX WPI; 2003-587286/55.
 DR
 DR N-PSDB; ADE87387.
 XX
 XX New pancreatic specific nucleic acid molecule or protein for diagnosing,
 PT staging, imaging, monitoring, preventing or treating pancreatic cancer or
 PT non-cancerous disease states of the pancreas.
 XX
 XX Claim 12; SEQ ID NO 510; 635pp; English.
 PS
 XX This invention relates to novel nucleic acids and proteins present in
 CC normal and neoplastic pancreatic cells. Pancreatic cancer is a common

CC cause of cancer death worldwide, therefore accurate methods of diagnosis
 CC and treatment are required. Compounds which modulate the proteins of the
 CC invention may have cytostatic activity and the protein and DNA sequences
 CC of the invention may be useful for the development of a vaccine or in
 CC gene therapy. The composition and methods are useful in diagnosing,
 CC staging, imaging, monitoring, preventing or treating pancreatic cancer
 CC and non-cancerous disease states of the pancreas. The present sequence is
 CC that of a human pancreatic protein of the invention.
 XX
 SQ Sequence 1161 AA;

Query Match 97.0%; Score 1149; DB 7; Length 1161;
 Best Local Similarity 93.6%; Pred. No. 6.9e-67;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GPPGPTGLPGERGSGRPGADGVAGPKGAGERGSPGPA 48
 DB 155 GPPGAGEGKRGARPEGPTGLPGERGSGRPGADGVAGPKGAGERGSPGPA 214
 QY 49 GPKSGPGEAGRPGEAGLPGAKLTGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 108
 DB 215 GPKSGPGEAGRPGEAGLPGAKLTGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 274
 QY 109 GVMGFPKGKGAAGSPKAGRGVPPGCVGPKAGDGEAGAGQPPGAGPAGERGEGGPA 168
 DB 275 GVMGFPKGKGAAGSPKAGRGVPPGCVGPKAGDGEAGAGQPPGAGPAGERGEGGPA 334
 QY 169 GSPGFQGLPGPAGPPGEAGKFGQVPGDLGAPGSPGAG 208
 DB 335 GSPGFQGLPGPAGPPGEAGKFGQVPGDLGAPGSPGAG 374

RESULT 10
 AAR89469
 ID AAR89469 standard; protein; 1169 AA.
 XX
 AC AAR89469;
 XX
 XX 01-OCT-1996 (first entry)
 DT
 XX Collagen/BMP-2B fusion protein.
 DE
 XX Bone morphogenic protein 2B; BMP-2B; collagen IA; osteogenesis;
 KW fusion protein.
 KW
 OS Synthetic.
 XX
 XX Key Location/Qualifiers
 FT Domain 1..1057
 FT /label= Collagen-IA
 FT /note= "collagen IA alpha-helical domain"
 FT Misc-difference 887
 FT /note= "unidentified amino acid"
 FT Misc-difference 890
 FT /note= "unidentified amino acid"
 FT Peptide 1058..1059
 FT /label= Linker_peptide
 FT Domain 1060..1169
 FT /label= BMP-2B
 FT /note= "human mature BMP-2B"
 FT
 XX CA2151547-A.
 PN
 XX 11-DEC-1995.
 PD
 XX 12-JUN-1995; 95CA-02151547.
 PF
 XX 10-JUN-1994; 94US-00259263.
 PR
 XX (USSU) US SURGICAL CORP.
 PA
 XX Gruskin EA, Espino P;
 PI
 XX

DR WPI; 1996-140144/15.
 DR N-PSDB; AAT16515.
 XX
 XX Chimaeric DNA encoding protein contg. extracellular matrix protein domain
 PT - and cellular regulatory factor domain, partic. useful as osteogenic
 PT agents, also related vectors, transformed cells and chimaeric proteins.
 XX
 XX Disclosure; Fig 5; 59pp; English.
 XX
 XX A fusion protein (AAR89469) comprises the alpha-helical region of human
 CC collagen I(a) linked to the human mature bone morphogenic protein 2B
 CC (BMP2B). It can be expressed in Escherichia coli transformants carrying a
 CC vector incorporating a chimeric gene (AAT16515) coding for the fusion.
 CC The BMP moiety induces osteogenesis, while the collagen moiety provides
 CC an integral substratum or scaffolding for the BMP and cells involved in
 CC reconstruction and growth. The fusion protein provides sustained release
 CC and delivery of BMP to a target tissue
 XX
 XX Sequence 1169 AA;
 Query Match 97.0%; Score 1149; DB 2; Length 1169;
 Best Local Similarity 93.6%; Pred. No. 7e-67;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GEPGTCLGPPQERGGPSRFPAGDVGAGPKGAGERSGPPA 48
 DB 297 GPPGAGEEGKRGARGEPGTGLPFPGERGGPSRFPAGDVGAGPKGAGERSGPPA 356
 QY 49 GPKGSPGEAGRPGEAGLPGAKGLTSGPSGDPGKTPGPGAGQDGRFPGPPPGARGQA 108
 DB 357 GPKGSPGEAGRPGEAGLPGAKGLTSGPSGDPGKTPGPGAGQDGRFPGPPPGARGQA 416
 QY 109 GVWGFPKGAGBFGKAGRGVPPGAVGPPAGKDGAGAGQPPGAGPAGRGEGGPA 168
 DB 417 GVWGFPKGAGBFGKAGRGVPPGAVGPPAGKDGAGAGQPPGAGPAGRGEGGPA 476
 QY 169 GSPGFQGLPGPAGPPGAGKPGEGQVPGDLGAPGSPGAG 208
 DB 477 GSPGFQGLPGPAGPPGAGKPGEGQVPGDLGAPGSPGAG 516
 RESULT 11
 AAY84537
 ID AAY84537 standard; protein; 1169 AA.
 AC AAY84537;
 XX
 XX 25-JUL-2000 (first entry)
 DT
 DE Amino acid sequence of a chimeric collagen 1 (alpha1)/BMP-2B protein.
 XX
 XX Extracellular matrix protein; self aggregation; hydroxylated proline;
 KW trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
 KW collagen; fibronectin; fibronectin; post translational hydroxylation;
 KW ss. bone morphogenic protein; BMP-2B; chimera.
 XX
 XX Homo sapiens.
 OS Unidentified.
 OS Chimeric.
 XX
 XX Key Location/Qualifiers
 FH
 FT Misc-difference 677
 FT /note= "Ala encoded by G"
 FT
 FT Misc-difference 887
 FT /note= "unspecified amino acid encoded by C"
 FT
 FT Misc-difference 890
 FT /note= "unspecified amino acid encoded by C"
 FT
 XX EP992586-A2.
 PN
 XX 12-APR-2000.
 PD
 XX 07-OCT-1999; 99EP-00119184.
 PF

XX
 PR
 XX
 PA (USSU) US SURGICAL CORP.
 XX
 PI Gruskin EA, Buechter DD, Zhang G, Connolly K;
 XX
 DR WPI; 2000-259138/23.
 DR N-PSDB; AAA12497.
 XX
 XX Production of extracellular matrix proteins containing 4-trans-
 PT hydroxyproline results in native self aggregating proteins, useful on
 PT medical implants.
 XX
 XX Claim 22; Fig 13; 260pp; English.
 XX
 XX The specification describes a method for producing an extracellular
 CC matrix protein or its fragment. The extracellular matrix protein is
 CC capable of self aggregating in a cell which does not ordinarily
 CC hydroxylated prolines. The method comprises optimising a nucleic acid
 CC sequence for expression in the cell by substitution of codons preferred
 CC by that cell for naturally occurring codons not preferred by the cell;
 CC incorporating the nucleic acid sequence into the cell; and contacting the
 CC cell with a hypertonic growth medium containing at least one amino acid,
 CC selected from the group consisting of trans-4-hydroxyproline and 3-
 CC hydroxyproline to allow at least one of the amino acids to be assimilated
 CC into the cell and incorporated into the extracellular matrix protein. The
 CC method may be used to make host cells assimilate and incorporate trans-4-
 CC hydroxyproline into proteins. This is especially useful in the
 CC recombinant production of proteins such as collagen, fibrinogen and
 CC fibronectin whose ability to self aggregate and produce functional
 CC proteins depends on the post translational hydroxylation of proline. The
 CC method is also useful in studying the structure and function of
 CC polypeptides which do not normally contain trans-4-hydroxyproline. The
 CC present sequence represents a chimeric collagen 1 (alpha1)/bone
 CC morphogenic protein-2B (bmp-2b) protein, which may be produced using the
 CC method of the invention
 XX
 SQ Sequence 1169 AA;
 Query Match 97.0%; Score 1149; DB 3; Length 1169;
 Best Local Similarity 93.6%; Pred. No. 7e-67;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GEPGTCLGPPGGERGGPSRFPAGDVGAGPKGAGERSGPPA 48
 DB 297 GPPGAGEEGKRGARGEPGTGLPFPGERGGPSRFPAGDVGAGPKGAGERSGPPA 356
 QY 49 GPKGSPGEAGRPGEAGLPGAKGLTSGPSGDPGKTPGPGAGQDGRFPGPPPGARGQA 108
 DB 357 GPKGSPGEAGRPGEAGLPGAKGLTSGPSGDPGKTPGPGAGQDGRFPGPPPGARGQA 416
 QY 109 GVWGFPKGAGBFGKAGRGVPPGAVGPPAGKDGAGAGQPPGAGPAGRGEGGPA 168
 DB 417 GVWGFPKGAGBFGKAGRGVPPGAVGPPAGKDGAGAGQPPGAGPAGRGEGGPA 476
 QY 169 GSPGFQGLPGPAGPPGAGKPGEGQVPGDLGAPGSPGAG 208
 DB 477 GSPGFQGLPGPAGPPGAGKPGEGQVPGDLGAPGSPGAG 516
 RESULT 12
 AAR89470
 ID AAR89470 standard; protein; 1171 AA.
 XX
 XX AAR89470;
 XX
 XX 01-OCT-1996 (first entry)
 DT
 XX Collagen/TGF-beta-1 fusion protein.
 DE
 XX Transforming growth factor; TGF-beta-1; collagen IA; osteogenesis;
 KW bone formation; tissue repair; fusion protein.
 KW

OS Synthetic.
 XX Key
 XX Domain
 FT Location/Qualifiers
 FT 1..1057
 FT /label= Collagen-IA
 FT /note= "collagen IA alpha-helical domain"
 FT Misc-difference 887
 FT /note= "unidentified amino acid"
 FT Misc-difference 890
 FT /note= "unidentified amino acid"
 FT Peptide
 FT 1058..1059
 FT /label= Linker_peptide
 FT Domain
 FT 1060..1171
 FT /label= TGF-beta-1
 FT /note= "human mature TGF-beta-1"
 XX
 XX CA2151547-A.
 PD 11-DEC-1995.
 XX
 XX 12-JUN-1995; 95CA-02151547.
 XX
 XX 10-JUN-1994; 94US-00259263.
 XX
 XX (USU) US SURGICAL CORP.
 XX
 XX Gruskin EA, Espino P;
 XX WPI; 1996-140144/15.
 DR N-PSDB; AAT16516.
 XX
 XX Chimaeric DNA encoding protein contg. extracellular matrix protein domain
 PT - and cellular regulatory factor domain, partic. useful as osteogenic
 PT agents, also related vectors, transformed cells and chimaeric proteins.
 XX
 PS Disclosure; Fig 6; 59pp; English.
 XX
 CC A fusion protein (AAR89470) comprises the alpha-helical region of human
 CC collagen I(a) linked to the human mature transforming growth factor beta-
 CC 1 (TGF-beta-1). It can be expressed in Escherichia coli transformants
 CC carrying a vector incorporating a chimeric gene (AAT16516) coding for the
 CC fusion. The TGF-beta- moiety increases efficacy of the body's normal soft
 CC tissue repair response and also induces osteogenesis. The collagen moiety
 CC provides an integral substratum or scaffolding for the TGF and cells
 CC involved in reconstruction and growth. The fusion protein provides
 CC sustained release and delivery of TGF-beta-1 to a target tissue
 XX
 SQ Sequence 1171 AA;
 Query Match 97.0%; Score 1149; DB 2; Length 1171;
 Best Local Similarity 93.6%; Pred. No. 7e-67;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GPPGPTGLPGERGSGRFGPCADGVAGPKGPGAGRGSGGPA 48
 Db 297 GPPGAGEGKRGARGEPGPTGLPGERGSGRFGPCADGVAGPKGPGAGRGSGGPA 356
 QY 49 GPKGSPGEAGRPGAGLPGAKGLTSGSPGPKGTGTPGAGQDGRPPGPPGARGQA 108
 Db 357 GPKGSPGEAGRPGAGLPGAKGLTSGSPGPKGTGTPGAGQDGRPPGPPGARGQA 416
 QY 109 GVMGFPKPKGNAEPGKAGRGVPGAVGPGAKDGEACGAQCPGPGAGRGQGGA 168
 Db 417 GVMGFPKPKGAAGPGKAGRGVPGAVGPGAKDGEACGAQCPGPGAGRGQGGA 476
 QY 169 GSPGFQGLPAGPPGAGKPGEGQVPGDLGAPGSPGAG 208
 Db 477 GSPGFQGLPAGPPGAGKPGEGQVPGDLGAPGSPGAG 516

RESULT 13
 AAY84538

ID AAY84538 standard; protein; 1171 AA.
 XX
 AC AAY84538;
 XX
 DT 25-JUL-2000 (first entry)
 XX
 DE A chimeric collagen 1 (alpha1)/TGF-beta1 protein.
 XX
 KW Extracellular matrix protein; self aggregation; hydroxylated proline;
 KW trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
 KW collagen; fibronectin; fibronectin; post translational hydroxylation;
 KW ss. transforming growth factor-beta1; TGF-beta1; chimera.
 XX
 XX Homo sapiens.
 OS Unidentified.
 OS Chimeric.
 XX
 XX Key Location/Qualifiers
 FT Misc-difference 858 /note= "Gly encoded by GCT"
 FT
 XX EP992586-A2.
 XX
 PD 12-APR-2000.
 XX
 XX 07-OCT-1999; 99EP-00119184.
 XX
 XX 09-OCT-1998; 98US-00169768.
 XX
 XX (USU) US SURGICAL CORP.
 XX
 XX Gruskin EA, Buechter DB, Zhang G, Connolly K;
 WPI; 2000-259138/23.
 DR N-PSDB; AAT12498.
 XX
 PT Production of extracellular matrix proteins containing 4-trans-
 PT hydroxyproline results in native self aggregating proteins, useful on
 PT medical implants.
 XX
 XX Claim 23; Fig 15; 260pp; English.
 XX
 CC The specification describes a method for producing an extracellular
 CC matrix protein or its fragment. The extracellular matrix protein is
 CC capable of self aggregating in a cell which does not ordinarily
 CC hydroxylated prolines. The method comprises optimising a nucleic acid
 CC sequence for expression in the cell by substitution of codons preferred
 CC by that cell for naturally occurring codons not preferred by the cell,
 CC incorporating the nucleic acid sequence into the cell; and contacting the
 CC cell with a hypertonic growth medium containing at least one amino acid,
 CC selected from the group consisting of trans-4-hydroxyproline and 3-
 CC hydroxyproline to allow at least one of the amino acids to be assimilated
 CC into the cell and incorporated into the extracellular matrix protein. The
 CC method may be used to make host cells assimilate and incorporate trans-4-
 CC hydroxyproline into proteins. This is especially useful in the
 CC recombinant production of proteins such as collagen, fibronectin and
 CC fibronectin whose ability to self aggregate and produce functional
 CC proteins depends on the post translational hydroxylation of proline. The
 CC method is also useful in studying the structure and function of
 CC polypeptides which do not normally contain trans-4-hydroxyproline. The
 CC present sequence represents chimeric collagen 1 (alpha1)/transforming
 CC growth factor-beta1 (TGF-beta1) protein, which may be produced using the
 CC method of the invention
 XX
 SQ Sequence 1171 AA;
 Query Match 97.0%; Score 1149; DB 3; Length 1171;
 Best Local Similarity 93.6%; Pred. No. 7e-67;
 Matches 206; Conservative 0; Mismatches 12; Indels 12; Gaps 1;
 QY 1 GPP-----GPPGPTGLPGERGSGRFGPCADGVAGPKGPGAGRGSGGPA 48
 Db 297 GPPGAGEGKRGARGEPGPTGLPGERGSGRFGPCADGVAGPKGPGAGRGSGGPA 356

QY	109	GVMGPPGKGAAGEFGKAGRGVFP	PGAVGPAGKDGAGAGQGGPPG	PAGPAGERGEGPA	168
Db	578	GVMGPPGKGAAGEFGKAGRGVFP	PGAVGPAGKDGAGAGQGGPPG	PAGPAGERGEGPA	637
QY	169	GSPGFQGLPGPAGPPGAGKPGEG	VGDLGAPGSGPAG	208	
Db	638	GSPGFQGLPGPAGPPGAGKPGEG	VGDLGAPGSGARG	677	

Search completed: September 24, 2004, 11:09:14
 Job time : 31.5883 secs

Sequence 1, Appl
Sequence 1, Appl
Sequence 79, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 49, Appl
Sequence 26, Appl
Sequence 8, Appl
Sequence 8, Appl
Sequence 18, Appl
Sequence 10, Appl
Sequence 1, Appl
Sequence 20, Appl
Sequence 5, Appl
Sequence 48, Appl
Sequence 48, Appl
Sequence 50, Appl
Sequence 22, Appl
Sequence 33, Appl
Sequence 12, Appl
Sequence 10, Appl
Sequence 10, Appl
Sequence 70, Appl
Sequence 35, Appl
Sequence 74, Appl
Sequence 22, Appl
Sequence 23, Appl
Sequence 4, Appl

16 2966 96.6 1464 16 US-10-357-851-1
17 2966 96.6 1464 16 US-10-358-024-1
18 2966 96.6 1464 16 US-10-734-564-79
19 2936 95.6 1463 15 US-10-402-089-2
20 2936 95.6 1463 15 US-10-402-072A-2
21 2857 93.1 822 15 US-10-342-331-49
22 2857 93.1 1453 16 US-10-458-091-26
23 2816.5 91.7 1449 15 US-10-402-089-8
24 2816.5 91.7 1449 15 US-10-402-072A-8
25 2791.5 90.9 1341 14 US-10-058-124-18
26 2226 72.5 1017 16 US-10-639-286-10
27 2209 72.0 1014 14 US-10-194-441A-1
28 2202 71.7 1418 14 US-10-058-124-20
29 2202 71.7 1418 16 US-10-458-091-5
30 2189 71.3 1014 14 US-10-194-441A-48
31 2189 71.3 1487 16 US-10-468-091-6
32 2156 70.2 595 15 US-10-342-331-48
33 2156 70.2 595 15 US-10-342-331-50
34 2045 66.6 500 12 US-10-232-175-22
35 2045 66.6 662 12 US-10-232-175-33
36 1924.5 62.7 1466 15 US-10-402-089-12
37 1924.5 62.7 1466 15 US-10-402-072A-12
38 1923 62.6 1366 15 US-10-402-089-10
39 1923 62.6 1366 15 US-10-402-072A-10
40 1921.5 62.6 1496 14 US-10-177-293-70
41 1921.5 62.6 1496 14 US-10-301-822-35
42 1921.5 62.6 1496 15 US-10-236-031B-74
43 1921.5 62.6 1496 16 US-10-458-091-22
44 1910 62.2 1497 16 US-10-458-091-23
45 1906.5 62.1 1466 15 US-10-402-089-4

ALIGNMENTS

RESULT 1
US-10-104-889-16
; Sequence 16, Application US/10104889
; Publication No. US20040086961A1
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BUCHTER, DOUGLAS
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/10/104,889
; FILING DATE: 22-Mar-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768
; FILING DATE: 09-OCT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEN, JEFFREY S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 228-8484
; TELEFAX: (516) 228-8516
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.
OM protein - protein search, using sw model
Run on: September 24, 2004, 11:07:31 ; Search time 67.1873 Seconds
(without alignments)
2603.588 Million cell updates/sec
Title: US-10-658-989A-4
Perfect score: 3070
Sequence: 1 GSECEGVRCGPPGPGAGA.....PGPSGDGPPGPPGPKGEG 544

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 1349238 seqs, 32158718 residues
Total number of hits satisfying chosen parameters: 1349238
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	ID	Description
1	2966	96.6	1057	16	US-10-104-889-16
2	2966	96.6	1057	16	US-10-104-889-20
3	2966	96.6	1107	16	US-10-104-889-11
4	2966	96.6	1169	16	US-10-104-889-6
5	2866	96.6	1171	16	US-10-104-889-8
6	2966	96.6	1388	16	US-10-104-889-10
7	2966	96.6	1461	16	US-10-458-091-25
8	2966	96.6	1464	12	US-09-918-715-261
9	2966	96.6	1464	14	US-10-060-036-159
10	2966	96.6	1464	14	US-10-171-311-36
11	2966	96.6	1464	14	US-10-216-705-21
12	2966	96.6	1464	14	US-10-149-352-2
13	2966	96.6	1464	14	US-10-177-293-65
14	2966	96.6	1464	14	US-10-301-822-28
15	2966	96.6	1464	15	US-10-291-265-243

```

;
; LENGTH: 1057 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID 1
US-10-104-889-16

```

```

Query Match      96.6%; Score 2966; DB 16; Length 1057;
Best Local Similarity 95.6%; Pred. No. 4.9e-146;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;

QY      1 GSEGGEGYRGSGPPGPAGAAAGADPCADGDEGCAKAGADACGACGAGFCFPGARGSPGE 60
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      201 GSEGGQGYRGSGPPGPAGAAAGADPCADGDEGCAKAGADACGAGFCFPGARGSPGQ 260

QY      61 GGGGPPGPKGSGSPGAPGSGKDTGAKGEGPPGVGVGGPPGAGECKPCARCEPPTGLP 120
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      261 GGGGPPGPKGNSGSPGAPGSGKDTGAKGEGPPGVGVGGPPGAGECKRGARCEPPTGLP 320

QY      121 GPPGRRGGPGSRGFPFGADGVAGPKPGPAGERGSGPCPAGPKGSPGEGAGRPGEAGLPGAKGLT 180
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      321 GPPGRRGGPGSRGFPFGADGVAGPKPGPAGERGSGPCPAGPKGSPGEGAGRPGEAGLPGAKGLT 380

QY      181 GSPGSGPGDGKTGPPGPAGEDRGPPGPGPAGAGEAGVWGFCPPKAGCEPCKAGERCVP 240
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      381 GSPGSGPGDGKTGPPGPAGEDRGPPGPGPAGAGEAGVWGFCPPKAGCEPCKAGERCVP 440

QY      241 GPPGAVGPAGKDXGAGAEAGPPGCPAGPAGERGEPGAGSPFGELGCPAGPGEAGKPGEE 300
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      441 GPPGAVGPAGKDXGAGAEAGPPGCPAGPAGERGEPGAGSPFGELGCPAGPGEAGKPGQE 500

QY      301 GVPGDLAGAPSGGARGEGPPGPERGVGPPGPPAGPGADGAPGDDGAKGDAGACAPGSE 360
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      501 GVPGDLAGAPSGGARGEGPPGPERGVGPPGPPAGPGRANGAPGNDGAKGDAGAPAPSGQ 560

QY      361 GAPGLEGMPGRRGAAGLUPGPKDRDGDAPKGAAGSGSGKDGVRGLTGPTGPPGAGAGDCK 420
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      561 GAPGLEGMPGRRGAAGLUPGPKDRDGDAPKGAAGSGSGKDGVRGLTGPTGPPGAGAGDCK 620

QY      421 GSGSGSGPAGPTGARGAGPDRGEPGPPGPAFAGPAGPPGADGCEPGAKGCEPDGAKGDAQPP 480
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      621 GSGSGSGPAGPTGARGAGPDRGEPGPPGPAFAGPAGPPGADGCEPGAKGCEPDGAKGDAQPP 680

QY      481 GPAGPAGPPGIGDVCAKAGKAGSGAPPGATGTFGMAAGRVGPPGSGDAGPPPPGPA 540
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      681 GPAGPAGPPGPTGNVAPKAGKAGSGAPPGATGTFGMAAGRVGPPGSGNAGPPPPGPA 740

QY      541 GREG 544
      |||||
Db      741 GREG 744
      |||||

```

RESULT 2
US-10-104-899-20
Sequence 20, Application US/10104899
Publication No. US2004008961A1
GENERAL INFORMATION:
APPLICANT: GRUSKIN, ELIOT A.
BUECHTER, DOUGLAS
BROKAW, JANE
ZHANG, GUANHUI
PLOELLA, DAVID
TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: DILWORTH & BARRESE
STREET: 333 EARLE OVINGTON BOULEVARD
CITY: UNIONDALE
STATE: NY
COUNTRY: U.S.A.
ZIP: 11553
COMPUTER READABLE FORM:

	Sequence 11, Application US/10104899	Query Match	96.6%; Score 2966; DB 16; Length 1107;
	Publication No. US20040086961A1	Best Local Similarity	95.6%; Pred. No. 5.1e-146;
	GENERAL INFORMATION:	Matches 520; Conservative	21; Mismatches 3; Indels 0; Gaps 0;
	APPLICANT: GRUSKIN, ELLIOT A.		
	BUECHTER, DOUGLAS		
	BROKAW, JANE		
	ZHANG, GUANGHUI		
	PAOLELLA, DAVID		
	TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES		
	NUMBER OF SEQUENCES: 50		
	CORRESPONDENCE ADDRESS:		
	ADDRESSEE: DILWORTH & BARRESE		
	STREET: 333 EARLE OWINGTON BOULEVARD		
	CITY: UNIONDALE		
	STATE: NY		
	COUNTRY: U.S.A.		
	ZIP: 11553		
	COMPUTER READABLE FORM:		
	MEDIUM TYPE: Floppy disk		
	COMPUTER: IBM PC compatible		
	OPERATING SYSTEM: PC-DOS/MS-DOS		
	SOFTWARE: Patentin Release #1.0, Version #1.30		
	CURRENT APPLICATION DATA:		
	APPLICATION NUMBER: US/10/104,889		
	FILING DATE: 22-Mar-2002		
	CLASSIFICATION: <Unknown>		
	PRIOR APPLICATION NUMBER: US/09/169,768		
	FILING DATE: 09-OCT-1998		
	ATTORNEY/AGENT INFORMATION:		
	NAME: STEEN, JEFFREY S		
	TELECOMMUNICATION INFORMATION:		
	TELEPHONE: (516) 228-8484		
	TELEFAX: (516) 228-8516		
	INFORMATION FOR SEQ ID NO: 11:		
	SEQUENCE CHARACTERISTICS:		
	LENGTH: 1107 amino acids		
	TYPE: amino acid		
	STRANDEDNESS: single		
	TOPOLOGY: unknown		
	MOLECULE TYPE: peptide		
	SEQUENCE DESCRIPTION: SEQ ID NO: 11:		
	US-10-104-889-11		
	Query Match	96.6%; Score 2966; DB 16; Length 1107;	
	Best Local Similarity	95.6%; Pred. No. 5.1e-146;	
	Matches 520; Conservative	21; Mismatches 3; Indels 0; Gaps 0;	
Qy	1 GSEGEVGRGPGPPAGAGPAGDPCADCEPKAKGADGAPGTAGAPFGARGSGPE 60		
Dd	201 GSEGPQGVRGEPGPPGAGAGPAGNPDQCPQAKANGAPGTAGAPFGARGSGPQ 260		
Qy	61 GGPGPPGPKDSGPCAGSKGDTCAKEGPFVGVEGPPGAPGEGKPGARGGPGTGLP 120		
Dd	261 GGPGPPGPKNGSFGAGPSKGDTCAKEGPFVGVEGPPGAPGEGKRGARGFPTGLP 320		
Qy	121 GPGERGGPSRFFGACGVAGPKGAPGERGSPGAPKSPGAGRPBAGLPGAKGLT 180		
Dd	321 GPFERGGPSRFFGACGVAGPKGAPGERGSPGAPKSPGAGRPBAGLPGAKGLT 380		
Qy	181 GSPPSGDPDKTPPGPAGEDRGPBPCCARGAEAGVMGPPGPKGAAGPPKAGERGVP 240		
Dd	381 GSPPSGDPDKTPPGPAGEDRGPBPCCARGAEAGVMGPPGPKGAAGPPKAGERGVP 440		
Qy	241 GPPGAVGAGKDGEAGAGPFPAPGAPGERGEEGAPSGPGFGLPGPAGPBGAKPGE 300		
Dd	441 GPPGAVGAGKDGEAGAGPFPAPGAPGERGEEGAPSGPGFGLPGPAGPBGAKPGE 500		
Qy	301 GVPDLGAPSPGARGEGPFGERGVEGPRGPPAGPDGAPGDGAKGAPGAPGSE 360		
Dd	501 GVPDLGAPSPGARGEGPFGERGVEGPRGPPAGPDGAPGDGAKGAPGAPGSE 560		
Qy	361 GAPGLEGMGERGAAGLPCKPKCDRDGAPKADGSPKDGVRGTTGPIGPPGAPAGDK 420		

[illegible]


```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/10/104,889
  FILING DATE: 22-Mar-2002
  CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US/09/169,768
  FILING DATE: 09-OCT-1998
ATTORNEY/AGENT INFORMATION:
  NAME: STEEN, JEFFREY S
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (516) 228-8484
    TELEFAX: (516) 228-8516
INFORMATION FOR SEQ ID NO: 10:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 1388 amino acids
    TYPE: amino acid
    STRANDEDNESS: single
    TOPOLOGY: unknown
  MOLECULE TYPE: peptide
  SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-10-104-889-10

Query Match      96.6%; Score 2966; DB 16; Length 1388;
Best Local Similarity 95.6%; Pred. No. 6.1e-146;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;

QY 1 GSEGPVGRGPPGPPGAGAGAGDPCADGEPGAKGADGAGTGTAGAGPPGARGSGPE 60
DB 201 GSEGPVGRGPPGPPGAGAGAGDPCADGEPGAKGADGAGTGTAGAGPPGARGSGPPQ 260
QY 61 GPGGPPGPKGDSGPPGAGPAGSGKDTGAKGEPVGVGPPGAGEGKPGARGPPGTGLP 120
DB 261 GPGGPPGPKGNSGPPGAGPAGSGKDTGAKGEPVGVGPPGAGEGKRGARGPPGTGLP 320
QY 121 GPPGPPGPPGPPGAGDAGVAGPKGAPGAGSGPPGAGPKSPGAPGPPGAGTGTAGT 180
DB 321 GPPGPPGPPGPPGAGDAGVAGPKGAPGAGSGPPGAGPKSPGAPGPPGAGTGTAGT 380
QY 181 GSPGPPGPPGPPGAGDAGVAGPKGAPGAGSGPPGAGPKSPGAPGPPGAGTGTAGT 240
DB 381 GSPGPPGPPGPPGAGDAGVAGPKGAPGAGSGPPGAGPKSPGAPGPPGAGTGTAGT 440
QY 241 GPPGPPGPPGPPGAGDAGVAGPKGAPGAGSGPPGAGPKSPGAPGPPGAGTGTAGT 300
DB 441 GPPGPPGPPGPPGAGDAGVAGPKGAPGAGSGPPGAGPKSPGAPGPPGAGTGTAGT 500
QY 301 GVPDGLGAPGPPGARGPPGPPGAGDAGVAGPKGAPGAGSGPPGAGPKSPGAPGPPG 360
DB 501 GVPDGLGAPGPPGARGPPGPPGAGDAGVAGPKGAPGAGSGPPGAGPKSPGAPGPPG 560
QY 361 GAPGLEMPGPPGPPGAGDAGVAGPKGAPGAGSGPPGAGPKSPGAPGPPGAGTGTAGT 420
DB 561 GAPGLEMPGPPGPPGAGDAGVAGPKGAPGAGSGPPGAGPKSPGAPGPPGAGTGTAGT 620
QY 421 GESGPPGPPGPPGAGDAGVAGPKGAPGAGSGPPGAGPKSPGAPGPPGAGTGTAGT 480
DB 621 GESGPPGPPGPPGAGDAGVAGPKGAPGAGSGPPGAGPKSPGAPGPPGAGTGTAGT 680
QY 481 GPAGPPGPPGPPGAGDAGVAGPKGAPGAGSGPPGAGPKSPGAGTGTAGTGTAGTGTAGT 540
DB 681 GPAGPPGPPGPPGAGDAGVAGPKGAPGAGSGPPGAGPKSPGAGTGTAGTGTAGTGTAGT 740
QY 541 GKEG 544
DB 741 GKEG 744

RESULT 7
US-10-468-091-25
; Sequence 25, Application US/10468091

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/10/104,899
  FILING DATE: 22-Mar-2002
  CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US/09/169,768
  FILING DATE: 09-OCT-1998
ATTORNEY/AGENT INFORMATION:
  NAME: STEEN, JEFFREY S
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (516) 228-8484
    TELEFAX: (516) 228-8516
INFORMATION FOR SEQ ID NO: 10:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 1388 amino acids
    TYPE: amino acid
    STRANDEDNESS: single
    TOPOLOGY: unknown
  MOLECULE TYPE: peptide
  SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-10-104-889-10

Query Match      96.6%; Score 2966; DB 16; Length 1388;
Best Local Similarity 95.6%; Pred. No. 6.1e-146;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;

QY 1 GSEGPVGRGPPGPPGAGAGAGDPCADGEPGAKGADGAGTGTAGAGPPGARGSGPE 60
DB 201 GSEGPVGRGPPGPPGAGAGAGDPCADGEPGAKGADGAGTGTAGAGPPGARGSGPPQ 260
QY 61 GPGGPPGPKGDSGPPGAGPAGSGKDTGAKGEPVGVGPPGAGEGKPGARGPPGTGLP 120
DB 261 GPGGPPGPKGNSGPPGAGPAGSGKDTGAKGEPVGVGPPGAGEGKRGARGPPGTGLP 320
QY 121 GPPGPPGPPGPPGAGDAGVAGPKGAPGAGSGPPGAGPKSPGAPGPPGAGTGTAGT 180
DB 321 GPPGPPGPPGPPGAGDAGVAGPKGAPGAGSGPPGAGPKSPGAPGPPGAGTGTAGT 380
QY 181 GSPGPPGPPGPPGAGDAGVAGPKGAPGAGSGPPGAGPKSPGAPGPPGAGTGTAGT 240
DB 381 GSPGPPGPPGPPGAGDAGVAGPKGAPGAGSGPPGAGPKSPGAPGPPGAGTGTAGT 440
QY 241 GPPGPPGPPGPPGAGDAGVAGPKGAPGAGSGPPGAGPKSPGAPGPPGAGTGTAGT 300
DB 441 GPPGPPGPPGPPGAGDAGVAGPKGAPGAGSGPPGAGPKSPGAPGPPGAGTGTAGT 500
QY 301 GVPDGLGAPGPPGARGPPGPPGAGDAGVAGPKGAPGAGSGPPGAGPKSPGAPGPPG 360
DB 501 GVPDGLGAPGPPGARGPPGPPGAGDAGVAGPKGAPGAGSGPPGAGPKSPGAPGPPG 560
QY 361 GAPGLEMPGPPGPPGAGDAGVAGPKGAPGAGSGPPGAGPKSPGAPGPPGAGTGTAGT 420
DB 561 GAPGLEMPGPPGPPGAGDAGVAGPKGAPGAGSGPPGAGPKSPGAPGPPGAGTGTAGT 620
QY 421 GESGPPGPPGPPGAGDAGVAGPKGAPGAGSGPPGAGPKSPGAPGPPGAGTGTAGT 480
DB 621 GESGPPGPPGPPGAGDAGVAGPKGAPGAGSGPPGAGPKSPGAPGPPGAGTGTAGT 680
QY 481 GPAGPPGPPGPPGAGDAGVAGPKGAPGAGSGPPGAGPKSPGAGTGTAGTGTAGTGTAGT 540
DB 681 GPAGPPGPPGPPGAGDAGVAGPKGAPGAGSGPPGAGPKSPGAGTGTAGTGTAGTGTAGT 740
QY 541 GKEG 544
DB 741 GKEG 744

RESULT 8
US-09-918-715-261
; Sequence 261, Application US/09918715
; Publication No. US20030017157A1
; GENERAL INFORMATION:
; APPLICANT: Brad St. Croix
; APPLICANT: Bert Vogelstein
; APPLICANT: Kenneth Kinzler
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS
; FILE REFERENCE: 1107.00134
```



```

; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 36
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-171-311-36

Query Match          96.6%; Score 2966; DB 14; Length 1464;
Best Local Similarity 95.6%; Pred. NO. 6.4e-146;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;

QY      1 GSEGGVGEVGEPPCPAGNAGPAGDGGEPGAKGADGAGIAGAPGFCARGPSGPE 60
DB      362 GSEGGVGEVGEPPCPAGNAGPAGDGGEPGAKGADGAGIAGAPGFCARGPSGPE 421
QY      61 GPGCGPPGPKGDSCEPCAPSGKGTGAKGPPGVGVGPPGCPAGEEGKPGARGEPFTGLP 120
DB      422 GPGCGPPGPKGNSGEPGAPGSKGDTGAKGPPGVGVGPPGCPAGEEGKPGARGEPFTGLP 481
QY      121 GPPGERGGSGRFFPGADGVAGPKGPAGRGSPGPPAGPKGSPCEAGRPOEAGLPGAKGLT 180
DB      482 GPPGERGGSGRFFPGADGVAGPKGPAGRGSPGPPAGPKGSPCEAGRPOEAGLPGAKGLT 541
QY      181 GSPGSGPPGDKTGPCCGPPAGGEGDRPPGPPGARGEGAGVMGFPGPKCAAGEPGKAGRGVYP 240
DB      542 GSPGSGPPGDKTGPCCGPPAGGEGDRPPGPPGARGEGAGVMGFPGPKCAAGEPGKAGRGVYP 601
QY      241 GPPGAVGPAGKCOGEAGCAAGPPGPPAGPAGRGEGEPGAGSPGFEGLPGAPGPEAGKPGEE 300
DB      602 GPPGAVGPAGKCOGEAGCAAGPPGPPAGPAGRGEGEPGAGSPGFEGLPGAPGPEAGKPGEE 661
QY      301 GVPGDILGAPSPGARGEPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 360
DB      662 GVPGDILGAPSPGARGEPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 721
QY      361 GAPPLEGMPGEEGAAGLPGPKGDRGAGKAGDAGSGPKDGVRLTGFPGPPGAPAGPDK 420
DB      722 GAPPLEGMPGEEGAAGLPGPKGDRGAGKAGDAGSGPKDGVRLTGFPGPPGAPAGPDK 781
QY      421 GESGSGGPPAGTGAAGAPGDRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 480
DB      782 GESGSGGPPAGTGAAGAPGDRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 841
QY      481 GPAGPAGPPGPIGDVAGPAGKAGGAGGAGPPGATGFCAGACRGVPPGPPSGDAGPPGPPGPA 540
DB      842 GPAGPAGPPGPIGDVAGPAGKAGGAGGAGPPGATGFCAGACRGVPPGPPSGDAGPPGPPGPA 901
QY      541 GKEG 544
DB      902 GKEG 905

RESULT 11
US-10-216-705-21
; Sequence 21, Application US/10216705
; Publication No. US2003009673A1
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics, S.A.
; TITLE OF INVENTION: Recombinant Collagens and Derived Proteins Produced by Plants, Me
; TITLE OF INVENTION: Obtaining Such and Their Uses
; FILE REFERENCE: 1149-3 DIV
; CURRENT APPLICATION NUMBER: US/10/216,705
; CURRENT FILING DATE: 2002-08-09

```

```

; PRIOR APPLICATION NUMBER: US 09/331,347
; PRIOR FILING DATE: 1999-08-17
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 21
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-216-705-21

```

```

Query Match      96.6%; Score 2366; DB 14; Length 1464;
Best Local Similarity 95.6%; Pred. No. 6.4e-146;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0
y
1 CSEGEVGRGPGPGPAGAGPACDPACDCEPGNKAGDAGPGTAGAGPFGAGPSPCE 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

[illegible]

```

RESULT 12
US-10-149-352-2
? Sequence 2, Application US/10149352
? Publication No. US20030105050A1
? GENERAL INFORMATION:
? APPLICANT: Beri, Rajinder
? TITLE OF INVENTION: ANTISENSE OLIGONUCLEOTIDES
? FILE REFERENCE: 06375-254US1
? CURRENT APPLICATION NUMBER: US/10/149,352
? CURRENT FILING DATE: 2003-06-10
? PRIOR APPLICATION NUMBER: PCT/GB06/04741
? PRIOR FILING DATE: 2000-12-12
? PRIOR APPLICATION NUMBER: GB 9929487.8
? PRIOR FILING DATE: 1999-12-15
? NUMBER OF SEQ ID NOS: 14
? SOFTWARE: Patentin Ver. 4.0
? SEQ ID NO 2
? LENGTH: 1464
?

```


GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF COLON CANCER
; FILE REFERENCE: MEM01-029P2RNM
; CURRENT APPLICATION NUMBER: US/10/301,822
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-301-822-28

Query Match 96.6%; Score 2966; DB 14; Length 1464;
Best Local Similarity 95.6%; Pred. No. 6.4e-146;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;
QY 1 GSEGEVGRGPPGPPAGAGAGDPCADCEPGKAGADAGAGTAGAGPPGARGPSGPE 60
Db 362 GSEGEVGRGPPGPPAGAGAGDPCADCEPGKAGADAGAGTAGAGPPGARGPSGPE 421
QY 61 GPGGPPGKDSGPPGAPGSKGDTCAKGEPPGVGVEGPPGAGGPPGARGPPGPTGLP 120
Db 422 GPGGPPGKDSGPPGAPGSKGDTCAKGEPPGVGVEGPPGAGGPPGARGPPGPTGLP 481
QY 121 GPPGCGGSGRGGPPGADGAGKGPAGGSGPPGAGKSGPPGAGPPGAGLPGAKGLT 180
Db 482 GPPGCGGSGRGGPPGADGAGKGPAGGSGPPGAGKSGPPGAGPPGAGLPGAKGLT 541
QY 181 GSPGSGPDGKTGPPGPPGAGEDGRPCPPGPPGARGGAGVMPGPPGAGGPPGARGGVP 240
Db 542 GSPGSGPDGKTGPPGPPGAGEDGRPCPPGPPGARGGAGVMPGPPGAGGPPGARGGVP 601
QY 241 GPPGAVGPKDGEAGAGGPPGPPGAGGPPGAGGPPGAGGPPGAGGPPGAGGPPG 300
Db 602 GPPGAVGPKDGEAGAGGPPGPPGAGGPPGAGGPPGAGGPPGAGGPPGAGGPPG 661
QY 301 GVPDGLGAPGSGARGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 360
Db 662 GVPDGLGAPGSGARGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 721
QY 361 GAGLEGMPGPPGAGLPGPKDGDGAGPKGADGSPGKDGVRGLTGPTGPPGAGAGDK 420
Db 722 GAGLEGMPGPPGAGLPGPKDGDGAGPKGADGSPGKDGVRGLTGPTGPPGAGAGDK 781
QY 421 GESGSGPAGTGAAGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 480
Db 782 GESGSGPAGTGAAGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 841
QY 481 GPAGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 540
Db 842 GPAGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 901
QY 541 GKEG 544
Db 902 GKEG 905

RESULT 15
US-10-291-265-243
; Sequence 243, Application US/10291265
; Publication No. US20030232054A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang et al
; TITLE OF INVENTION: NO. US20030232054A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-017 (785)
; CURRENT APPLICATION NUMBER: US/10/291,265
; CURRENT FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 243
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-291-265-243

Query Match 96.6%; Score 2966; DB 15; Length 1464;
Best Local Similarity 95.6%; Pred. No. 6.4e-146;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;
QY 1 GSEGEVGRGPPGPPAGAGAGDPCADCEPGKAGADAGAGTAGAGPPGARGPSGPE 60
Db 362 GSEGEVGRGPPGPPAGAGAGDPCADCEPGKAGADAGAGTAGAGPPGARGPSGPE 421
QY 61 GPGGPPGKDSGPPGAPGSKGDTCAKGEPPGVGVEGPPGAGGPPGARGPPGPTGLP 120
Db 422 GPGGPPGKDSGPPGAPGSKGDTCAKGEPPGVGVEGPPGAGGPPGARGPPGPTGLP 481
QY 121 GPPGCGGSGRGGPPGADGAGKGPAGGSGPPGAGKSGPPGAGPPGAGLPGAKGLT 180
Db 482 GPPGCGGSGRGGPPGADGAGKGPAGGSGPPGAGKSGPPGAGPPGAGLPGAKGLT 541
QY 181 GSPGSGPDGKTGPPGPPGAGEDGRPCPPGPPGARGGAGVMPGPPGAGGPPGARGGVP 240
Db 542 GSPGSGPDGKTGPPGPPGAGEDGRPCPPGPPGARGGAGVMPGPPGAGGPPGARGGVP 601
QY 241 GPPGAVGPKDGEAGAGGPPGPPGAGGPPGAGGPPGAGGPPGAGGPPGAGGPPG 300
Db 602 GPPGAVGPKDGEAGAGGPPGPPGAGGPPGAGGPPGAGGPPGAGGPPGAGGPPG 661
QY 301 GVPDGLGAPGSGARGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 360
Db 662 GVPDGLGAPGSGARGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 721
QY 361 GAGLEGMPGPPGAGLPGPKDGDGAGPKGADGSPGKDGVRGLTGPTGPPGAGAGDK 420
Db 722 GAGLEGMPGPPGAGLPGPKDGDGAGPKGADGSPGKDGVRGLTGPTGPPGAGAGDK 781
QY 421 GESGSGPAGTGAAGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 480
Db 782 GESGSGPAGTGAAGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 841
QY 481 GPAGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 540
Db 842 GPAGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 901
QY 541 GKEG 544
Db 902 GKEG 905

Fri Sep 24 11:45:39 2004

us-10-658-989a-4.rapb

Page 10

Search completed: September 24, 2004, 11:13:25
Job time : 69.1873 secs

- J. Biol. Chem. 260, 2315-2320, 1985
 A>Title: Fine structural analysis of the human pro-alpha 1 (I) collagen gene. Promoter s
 A:Reference number: 155237; MUID:85130970; PMID:2857713
 A:Accession: 155237
 A>Status: translation not shown; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-34 <CH2>
 A:Cross-references: GB:M10627; NID:G180383; PIDN:AAA51992.1; PID:G553226
 R:Wirtz, M.K.; Keene, D.R.; Hori, H.; Glanville, R.W.; Steinmann, B.; Rao, V.H.; Hollist
 J. Biol. Chem. 265, 6312-6317, 1990
 A>Title: In vivo and in vitro noncovalent association of excised alpha(I) amino-termina
 lone, type VII.
 A:Reference number: A35233; MUID:90202908; PMID:2318955
 A:Accession: A35233
 A:Molecule type: protein
 A:Residues: 33-52 <WTR>
 A>Note: This propeptide fragment remained non-covalently bound to a defective, uncleaved
 E:Wail, D.; d'Alessio, M.; Ramirez, F.; de Wet, W.; Cole, W.G.; Chan, D.; Bateman, J.F.
 EMBO J. 8, 1705-1710, 1989
 A>Title: A base substitution in the exon of a collagen gene causes alternative splicing
 A:Reference number: S09400; MUID:89356643; PMID:2767050
 A:Accession: S09400
 A:Molecule type: mRNA
 A:Residues: 156-183 <WEI>
 R:Click, E.M.; Bornstein, P.
 Biochemistry 9, 4699-4706, 1970
 A>Title: Isolation and characterization of the cyanogen bromide peptides from the alpha
 A:Reference number: A90567; MUID:71038625; PMID:5529814
 A:Contents: CNBr0-1, CNBr2, CNBr4, CNBr5
 A:Accession: B90567
 A:Molecule type: protein
 A:Residues: 182-198, 'Z', 200-201, 'Z', 203-206, 'Z', 208-209, 'Z', 211-228, 'B', 230, 'BB', 233, 'Z'
 A:Experimental source: skin
 A>Note: Evidence for 170-allysine
 R:Baetge, B.; Norbom, H.; Diebold, J.; Lehmann, H.; Bodo, M.; Deutzmann, R.; Mueller, F.
 Eur. J. Biochem. 192, 153-159, 1990
 A>Title: A critical crosslink region in human-bone-derived collagen type I. Specific cle
 A:Reference number: S11372; MUID:90382436; PMID:2169412
 A:Accession: S11372
 A:Molecule type: protein
 A:Residues: 175-187; 274-287, 'P', 289 <BAE>
 A>Note: sequence of collagen alpha 1(S)(I) isolated from bone after pepsin digestion
 R:Deak, S.B.; Scholz, P.M.; Amenta, P.S.; Constantinou, C.D.; Levi-Minzi, S.A.; Gonzalez
 J. Biol. Chem. 266, 21827-21832, 1991
 A>Title: The substitution of arginine for glycine 85 of the alpha 1(I) procollagen chain
 cooperative melting of intact type I collagen.
 A:Reference number: 155342; MUID:92042092; PMID:1718984
 A:Accession: 155342
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 258-268; 1347-1357 <DEA>
 A:Cross-references: GB:S67495; NID:9239007; PIDN:AAB20350.1; PID:G239008
 A>Note: sequences from the 5' and 3' ends only are shown; mutant sequence 263-Arg report
 R:Morgan, P.H.; Jacobs, H.G.; Segrest, J.P.; Cunningham, L.W.
 J. Biol. Chem. 245, 5042-5048, 1970
 A>Title: Comparative study of glycopeptides derived from selected vertebrate collagens.
 A:Reference number: A92069; MUID:71001508; PMID:4319110
 A:Accession: A92069
 A:Molecule type: protein
 A:Residues: 263-268 <MOP>
 A:Experimental source: skin
 A>Note: attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose to 5-hydroxylysine
 R:Labhard, M.E.; Hollister, D.W.
 Matrix 10, 124-130, 1990
 A>Title: Segmental amplification of the entire helical and telopeptide regions of the cD
 A:Reference number: S15989; MUID:90326017; PMID:2374517
 A:Accession: S15989
 A:Molecule type: mRNA
 A:Residues: 281-302; 402-420; 823-843; 925-944; 1026-1045; 1143-1162 <IAB>
 R:Wirtz, M.K.; Rao, V.H.; Glanville, R.W.; Labhard, M.E.; Pretorius, P.J.; de Vries, W.N.
 Connect. Tissue Res. 29, 1-11, 1993
 A>Title: A cysteine for glycine substitution at position 175 in an alpha 1 (I) chain of
 A:Reference number: 152905; MUID:93339042; PMID:8339541
 A:Accession: 152905
- A:Accession: 152905
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 342-352, 'C', 354-359 <W12>
 A:Cross-references: GB:S64717; NID:G408195; PIDN:AAB27677.1; PID:G408196
 A>Note: mutant sequence from patient with osteogenesis imperfecta
 R:Bernard, M.P.; Chu, M.L.; Myers, J.C.; Ramirez, F.; Eikenberry, E.F.; Prockop, D.J.
 Biochemistry 22, 5213-5223, 1983
 A>Title: Nucleotide sequences of complementary deoxyribonucleic acids for the proalpha 1
 A:Reference number: A90476; MUID:84080385; PMID:6689127
 A:Accession: A90476
 A:Molecule type: mRNA
 A:Residues: 425-1250, 'X', 1252-1328, 'S', 1330-1390, 'X', 1392-1464 <BER>
 A:Cross-references: GB:K01238; NID:G180391; PIDN:AAA51995.1; PID:G180392
 A>Note: sequence partially completed for missing nucleotides by A29439
 R:Chu, M.L.; Gargiulo, V.; Williams, C.J.; Ramirez, F.
 J. Biol. Chem. 260, 691-694, 1985
 A>Title: Multixon deletion in an osteogenesis imperfecta variant with increased type III
 A:Reference number: A22161; MUID:85104934; PMID:2981843
 A:Accession: A22161
 A:Molecule type: DNA
 A:Residues: 472-594, 'R', 596-607 <CH3>
 A:Cross-references: GB:K03178; GB:K03179; NID:G179612; NID:G179613; PIDN:AAA51847.1; PID:
 R:Wallis, G.A.; Starman, B.J.; Zinn, A.B.; Byers, P.H.
 Am. J. Hum. Genet. 46, 1034-1040, 1990
 A>Title: Variable expression of osteogenesis imperfecta in a nuclear family is explained
 A:Reference number: A35336; MUID:90252792; PMID:2339700
 A:Accession: A35336
 A:Molecule type: mRNA
 A:Residues: 710-720, 'E', 722-737, 'E', 739-745 <WAL>
 A>Note: the authors translated the codons CAG for 721 and CGT for 738 as Glu
 R:Forlino, A.; Zolezzi, F.; Valli, M.; Pignatti, P.F.; Cetta, G.; Brunelli, P.C.; Mottes,
 Hum. Mol. Genet. 3, 2201-2206, 1994
 A>Title: Severe (type III) osteogenesis imperfecta due to glycine substitutions in the ce
 A:Reference number: 154365; MUID:95187161; PMID:7881420
 A:Accession: 154365
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 748-766, 'S', 768-781 <FOR>
 A:Cross-references: GB:L47667; NID:G1009093; PIDN:AAB59576.1; PID:G1009094
 R:Chessler, S.D.; Wallis, G.A.; Byers, P.H.
 J. Biol. Chem. 268, 18218-18225, 1993
 A>Title: Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I) chain of ty
 A:Reference number: A47426; MUID:93352646; PMID:8349697
 A:Accession: A47426
 A:Molecule type: mRNA
 A:Residues: 1179-1276, 'H', 1278-1336, 1339-1387, 'R', 1389-1464 <CHE>
 A:Cross-references: GB:S64596; NID:G407589; PIDN:AAB27856.1; PID:G407590
 A>Note: sequence extracted from NCBI backbone (NCBIN:136444, NCBI:136445)
 A>Note: does not represent an experimentally determined sequence but three different muta
 A:Accession: B47426
 A:Molecule type: mRNA
 A:Residues: 1179-1464 <CH4>
 A:Experimental source: normal dermal fibroblast culture
 A:Accession: C47426
 A:Molecule type: mRNA
 A:Residues: 1179-1276, 'H', 1278-1464 <CH5>
 A:Experimental source: fetal cell 86-237
 A:Accession: D47426
 A:Molecule type: mRNA
 A:Residues: 1179-1336, 1339-1464 <CH6>
 A:Experimental source: fetal cell 86-146
 A:Accession: E47426
 A:Molecule type: mRNA
 A:Residues: 1179-1387, 'R', 1389-1464 <CH7>
 A:Experimental source: fetal cell 88-251
 R:Comol, D.H.; Apone, S.; Eyre, D.R.; Starman, B.J.; Andreassen, P.; Charbonneau, H.; Nicl
 J. Biol. Chem. 263, 14605-14607, 1988
 A>Title: Substitution of Cysteine for Glycine within the Carboxyl-terminal Telopeptide of
 A:Reference number: 155269; MUID:89008319; PMID:3170557
 A:Accession: 155269

A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1187-1194, 'C', 1196-1220 <CO>
A/Cross-references: GB:M3213; NID:g340842; PIDN:AA59363.1; PID:g495622
A/Note: mutant sequence from a patient with mild osteogenesis imperfecta
R:Maekelae, J.K.; Raasina, M.; Vitta, A.; Vuorio, E.
Nucleic Acids Res. 16, 349, 1988
A/Title: Human pro-alpha-1(I) collagen: cDNA sequence for the C-propeptide domain.

Query Match 96.6%; Score 2966; DB 1; Length 1464;
Best Local Similarity 95.6%; Pred. No. 5e-137;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;

QY 1 GSEGEVGRGPGPPGAGAGPAGDPCADGEPGAKGADGAPGIACAGPFCARGPSPE 60
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 60
QY 362 GSEGPQGVGRGPGPPGAGAGPAGNPGADQPGAKGANGAPGIAGAPFCFPGARGPSQ 421
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 421
QY 422 GPGGPPGPKGNSGEPGAPGSKDGTGAKGEPGVQGPAGGEGKGGARCEPPTGLP 120
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 120
QY 121 GPPERGGPGSRGPPGADGVAGPKGPAGERSGPGAPGKSPGEGAGRCEAGLPGAKGLT 180
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 180
QY 482 GPPERGGPGSRGPPGADGVAGPKGPAGERSGPGAPGKSPGEGAGRCEAGLPGAKGLT 541
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 541
QY 181 GSPGSPDPDKTGPFPAGEGDRPGPPGPGARGAGVMGPPGKGAAGEPGKAGERGVP 240
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 240
QY 542 GSPGSPDPDKTGPFPAGAGQGRPGPPGARGAGVMGPPGKGAAGEPGKAGERGVP 601
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 601
QY 241 GPPGAVGPAKGAGAGAGPPGPPGAPGAGERGEAGSGPGFEGLPGPAGPCEAGKPGEE 300
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 300
QY 602 GPPGAVGPAKGAGAGAGPPGPPGAPGAGERGEAGSGPGFEGLPGPAGPCEAGKPGEE 661
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 661
QY 301 GVPGLGAPGSPGARGEPGFGERGVEGPPGAPGPGADGAPGDDGAKGDAGAPGAPGSE 360
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 360
QY 662 GVPGLGAPGSPGARGERGFGERGVEGPPGAPGPGADGAPGDDGAKGDAGAPGAPG 721
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 721
QY 361 GAPLEGMPGSRGAAGLPFGPKGDRGADGPKGADGSPGKDGVRGLTGPFGPPGAPAGD 420
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 420
QY 722 GAPGLQMPGSRGAAGLPFGPKGDRGADGPKGADGSPGKDGVRGLTGPFGPPGAPAGD 781
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 781
QY 421 GESGSPGAGTGAAGAPCDRGEPGPPGAPGAPGADGEPGAKGEGFDGAKGDAGPP 480
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 480
QY 782 GESGSPGAGTGAAGAPCDRGEPGPPGAPGAPGADGEPGAKGEGFDGAKGDAGPP 841
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 841
QY 481 GPAGPAGPPGPIGVGAPGAKGAGSAGPPGATGPPGAAGRVGPPGSGDAGPPGPPGA 540
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 540
QY 842 GPAGPAGPPGPIGVGAPGAKGAGSAGPPGATGPPGAAGRVGPPGSGNAGPPGPPGA 901
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 901
QY 541 GKGG 544
Db |||||
QY 902 GKGG 905

RESULT 2
S21626
collagen alpha 1(I) chain precursor - mouse
C/Species: Mus musculus (house mouse)
C/Date: 13-Jan-1995 #sequence revision 25-Apr-1997 #text change 13-Aug-1999
C/Accession: S57243; S16374; A23982; I49559; I49557; S39789; I48300; S21626
R:Li, S.W.; Khillan, J.; Prockop, D.J.
Matrix Biol. 14, 593-595, 1994
A/Title: The complete cDNA coding sequence for the mouse pro-alpha-1(I) chain of type I
A/Reference number: S57243
A/Accession: S57243
A:Molecule type: mRNA
A:Residues: 1-1453
A/Cross-references: EMBL:U08020; NID:g470673; PIDN:AAA88912.1; PID:g470674
R:Maeseranta, M.; Tomari, D.; de Crombrugne, B.; Vuorio, E.
Biochim. Biophys. Acta 1089, 241-243, 1991
A/Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.
A/Reference number: S16176; MUID:91274355; PMID:2054384
A/Accession: S16374

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1442-1453 <MRT>
A:Cross-references: EMBL:X57981; NID:G50484; PIDN:CRAA41046.1; PID:G50485
R:French, B.T.; Lee, W.H.; Maul, G.G.
Gene 39, 311-312, 1985
A:Title: Nucleotide sequence of a cDNA clone for mouse proalpha1(I) collagen protein.
A:Reference number: A23982; MUID:86137403; PMID:3841523
A:Accession: A23982
A:Molecule type: mRNA
A:Residues: 518-1128 <FRE>
A:Cross-references: GB:M14423; NID:G192261; PIDN:AAA37333.1; PID:G192262
R:Monson, J.M.; Friedman, J.; McCarthy, B.J.
Mol. Cell. Biol. 2, 1362-1371, 1982
A:Title: DNA sequence analysis of a mouse pro-alpha-1(I) procollagen gene: Evidence for
A:Reference number: I49559; MUID:83141374; PMID:6298597
A:Accession: I49559
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 735-1130 <RES>
A:Cross-references: GB:M17491; NID:G192263; PIDN:AAA37334.1; PID:G192264
R:Harbers, K.; Kuehn, M.; Deltus, H.; Jaenisch, R.
Proc. Natl. Acad. Sci. U.S.A. 81, 1504-1508, 1984
A:Title: Insertion of retrovirus into the first intron of alpha1(I) collagen gene leads to
A:Reference number: I49557; MUID:84170331; PMID:6324198
A:Accession: I49557
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-25 <RE2>
R:Fenton, S.P.; Lamande, S.R.; Hannagan, M.; Stacey, A.; Jaenisch, R.; Bateman, J.F.
Biochim. Biophys. Acta 1216, 469-474, 1993
A:Title: Genomic sequence of mouse COL1A1 encoding the collagen propeptides.
A:Reference number: S39789; MUID:94092741; PMID:8269229
A:Accession: S39789
A:Molecule type: DNA
A:Residues: 1-80, 'E', 82-105, 'D', 107-185, 1031-1201, 'G', 1203-1218, 'E', 1220-1221, 'T', 1223-11
R:Rhodes, K.; Rippe, R.A.; Umezawa, A.; Nehls, M.; Brenner, D.A.; Breindl, M.
Mol. Cell. Biol. 14, 5950-5960, 1994
A:Title: DNA methylation represses the murine alpha 1(I) collagen promoter by an indirect
A:Reference number: I48300; MUID:94344105; PMID:8065328
A:Accession: I48300
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-80, 'E', 82-105, 'D', 107-147 <REF>
A:Cross-references: EMBL:X54876; NID:G50486; PIDN:CRAA38657.1; PID:G50487
C:Genetics:
A:Gene: COL1A1
A:Introns: 770/3; 789/3; 806/3; 842/3; 860/3; 878/3; 932/3; 968/3; 1004/3; 1022/3; 1058/3
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: coiled coil; extracellular matrix; glycoprotein; heterotrimer; triple helix
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-151/Domain: amino-terminal propeptide #status predicted <PRO>
F:30-89/Domain: von Willebrand factor type C repeat homology <VWC>
F:152-1453/Product: collagen alpha 1(I) chain #status predicted <MAT>
F:15224-1453/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

	Query Match	93.1%;	Score 2857;	DB 2;	Length 1453;	
	Best Local Similarity	91.4%;	Pred. No. 9.5e-132;			
	Matches 497;	Conservative 28;	Mismatches 19;	Indels 0;	Gaps 0;	
<hr/>						
Qy	1	GSEPGVGRGEPGP	PAGAAAGPAGDPCADGE	PGAKGADGAPGIAGAPF	PGARGSPGE	60
		: : : : :	: : : : :	: : : : :	: : : : :	
		: : : : :	: : : : :	: : : : :	: : : : :	
Dd	351	GSEPGVGRGEPGP	PAGAAGPAGNPGADQP	CAKGANGAPGIAGAPF	PGARGSPGQ	410
		: : : : :	: : : : :	: : : : :	: : : : :	
		: : : : :	: : : : :	: : : : :	: : : : :	
Qy	61	GPQGPPPKDSD	GGPPGAPSGKDTGAKGE	PPGVGEGPPGAGEGK	PGARPEPTGLP	120
		: : : : :	: : : : :	: : : : :	: : : : :	
		: : : : :	: : : : :	: : : : :	: : : : :	
Dd	411	GPSPPPFGKNS	GFPGPNKGTGAKGE	PGATGYCGPPPGAGEG	KTGARGEPGSGLP	470
		: : : : :	: : : : :	: : : : :	: : : : :	
		: : : : :	: : : : :	: : : : :	: : : : :	
Qy	121	GPPEERGPGSR	GGPPGADVAGPKPGAGE	RSGCPGKSPGCEAGL	PGNAKGLT	180
		: : : : :	: : : : :	: : : : :	: : : : :	
		: : : : :	: : : : :	: : : : :	: : : : :	
Dd	471	GPPEERGPGSR	GGPPGADVAGPKPGAGE	RSGCPGKSPGCEAGL	PGNAKGLT	530
		: : : : :	: : : : :	: : : : :	: : : : :	
		: : : : :	: : : : :	: : : : :	: : : : :	

181 GSPGSGPDKTGTGGPAGDGRGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 240
 531 GSPGSGPDKTGTGGPAGDGRGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 590
 241 GPPGAVGPKDDEAGAGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 300
 591 GPPGAVGPKDDEAGAGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 650
 301 GVPDGLGAPGSGARGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 360
 651 GVPDGLGAPGSGARGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 710
 361 GARGLEKMPGSGARGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 420
 711 GARGLEKMPGSGARGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 770
 421 GESGSPGAGTARGAGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 480
 771 GEAGSPGAGTARGAGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 830
 481 GPAGPAGPPGPPGARGAGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 540
 831 GPAGPAGPPGPPGARGAGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 890
 541 GKEG 544
 891 GKEG 894

RESULT 3
 CGCHLS
 collagen alpha 1(I) chain - chicken (tentative sequence) (fragments)
 C:Species: Gallus gallus (chicken)
 C:Date: 12-Aug-1981 #sequence_revision 06-Jul-1982 #text_change 31-Mar-2000
 C:Accession: A90458; A90181; A02857
 R:Highberger, J.H.; Corbett, C.; Dixit, S.N.; Yu, W.; Seyer, J.M.; Kang, A.H.; Gross, J.
 Biochemistry 21, 2048-2055, 1982
 A:Title: Amino acid sequence of chick skin collagen alpha1(I)-C88 and the complete primary structure of the alpha1(I)-C88
 A:Reference number: A90458; MUID:82231995; PMID:7093229
 A:Accession: A90458
 A:Molecule type: protein
 A:Residues: 1-1036 <HRG>
 A:Experimental source: skin
 A:Note: This is the latest in a series of papers from these workers elucidating the sequence of the alpha1(I) chain of chick skin collagen.
 R:Eye, D.R.; Glincher, M.J.
 Biochem. Biophys. Res. Commun. 48, 720-726, 1972
 A:Title: Evidence for a previously undetected sequence at the carboxyterminus of the alpha1(I) chain of chick skin collagen
 A:Reference number: A90181; MUID:72243016; PMID:5047697
 A:Accession: A90181
 A:Molecule type: protein
 A:Residues: 1037-1042 <EVR>
 A:Experimental source: skin
 A:Note: residues 1037-1042 above correspond to the carboxyl end of the protein
 C:Comment: Lysines at positions 1037, 700, 934, and 946 above may be hydroxylated in some species.
 C:Comment: Most of the prolines at the third position of the tripeptide repeating unit (Pro-Gly-Hyp) are hydroxylated.
 C:Comment: Pro-1002 is the only 3-hydroxyproline and the only hydroxylated proline in the collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology; C:Superfamily: coiled coil; extracellular matrix; glycoprotein; pyroglyutamic acid; trimer; F:1/Modified site: pyroglutamic acid (Gln) #status experimental

Query Match 90.4%; Score 2775; DB 1; Length 1042;
 Best Local Similarity 89.0%; Pred. No. 6.9e-128;
 Matches 484; Conservative 28; Mismatches 32; Indels 0; Gaps 0;

1 GSEGEVGRGPPGPPGARGAGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 60
 200 GSEGEVGRGPPGPPGARGAGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 259
 61 GPGGPPGPKDDEAGAGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 120
 260 GPGGPPGPKDDEAGAGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 319
 121 GPPGAVGPKDDEAGAGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 180

320 GPAGERGAPSGRPPGADGIAGPKGPPGERGSPGAVGFKGSPGEAGPPGAGLPKAGLT 379
 181 GSPGSGPDKTGTGGPAGDGRGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 240
 380 GSPGSGPDKTGTGGPAGDGRGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 439
 241 GPPGAVGPKDDEAGAGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 300
 440 GPPGAVGPKDDEAGAGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 499
 301 GVPDGLGAPGSGARGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 360
 500 GVPDGLGAPGSGARGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 559
 361 GARGLEKMPGSGARGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 420
 560 GARGLEKMPGSGARGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 619
 421 GESGSPGAGTARGAGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 480
 620 GEAGSPGAGTARGAGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 679
 481 GPAGPAGPPGPPGARGAGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 540
 680 GPAGPAGPPGPPGARGAGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 739
 541 GKEG 544
 740 GKZG 743

RESULT 4
 T45467
 collagen alpha 1(II) chain precursor [imported] - horse
 N:Alternate names: type II collagen
 C:Species: Equus caballus (domestic horse)
 C:Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 04-Mar-2000
 C:Accession: T45467
 R:Richardson, D.W.; Dodge, G.R.
 submitted to the EMBL Data Library, June 1996
 A:Description: Cloning of equine type II collagen and modulation of its expression in eq
 A:Reference number: Z22977
 A:Accession: T45467
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1418 <RIC>
 A:Cross-references: EMBL:U62528; PIDN:AAB05773.1
 C:Superfamily: collagen alpha 1(II) chain; fibrillar collagen carboxyl-terminal homology;

Query Match 72.3%; Score 2220; DB 2; Length 1418;
 Best Local Similarity 72.6%; Pred. No. 6.5e-101;
 Matches 395; Conservative 37; Mismatches 112; Indels 0; Gaps 0;

1 GSEGEVGRGPPGPPGARGAGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 60
 315 GSEGEVGRGPPGPPGARGAGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 374
 61 GPGGPPGPKDDEAGAGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 120
 375 GATGGLGPKDDEAGAGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 434
 121 GPPGAVGPKDDEAGAGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 180
 435 GPPGAVGPKDDEAGAGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 494
 181 GSPGSGPDKTGTGGPAGDGRGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 240
 495 GPPGAVGPKDDEAGAGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 554
 241 GPPGAVGPKDDEAGAGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 300
 555 GAPGLRGLPKDDEAGAGPPGPPGARGAGVGMFPKGAAGEPKAGRGV 614

A;Cross-references: EMBL:X16711; NID:g30040; PIDN:CAA34683.1; PID:g30041
A;Note: alternative splice form 1
R;Ryan, M.C.; Sandell, L.J.
J. Biol. Chem. 265, 10334-10339, 1990
A;Title: Differential expression of a cysteine-rich domain in the amino-terminal propept
A;Reference number: A35428; MUID:90285153; PMID:2355003
A;Accession: A35428
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 27-81,'L',83-103 <RYA2>
A;Note: alternative splice form 2; splicing appears to be under developmental regulation
R;Su, M.W.; Benson-Chanda, V.; Vissing, H.; Ramirez, F.
Genomics 4, 438-441, 1989
A;Title: Organization of the exons coding for Pro alpha-1(II) collagen N-propeptide conf
A;Reference number: A30147; MUID:89233138; PMID:2714801
A;Accession: A30147
A;Molecule type: DNA
A;Residues: 104-157,'P',159-236 <SUM>
A;Cross-references: GB:J03065; GB:M23660; GB:M25655; GB:M25656; GB:M25730; GB:M32168; GB:M32169; GB:M32170; GB:M32171; GB:M32172; GB:M32173; GB:M32174; GB:M32175; GB:M32176; GB:M32177; GB:M32178; GB:M32179; GB:M32180; GB:M32181; GB:M32182; GB:M32183; GB:M32184; GB:M32185; GB:M32186; GB:M32187; GB:M32188; GB:M32189; GB:M32190; GB:M32191; GB:M32192; GB:M32193; GB:M32194; GB:M32195; GB:M32196; GB:M32197; GB:M32198; GB:M32199; GB:M32200; GB:M32201; GB:M32202; GB:M32203; GB:M32204; GB:M32205; GB:M32206; GB:M32207; GB:M32208; GB:M32209; GB:M32210; GB:M32211; GB:M32212; GB:M32213; GB:M32214; GB:M32215; GB:M32216; GB:M32217; GB:M32218; GB:M32219; GB:M32220; GB:M32221; GB:M32222; GB:M32223; GB:M32224; GB:M32225; GB:M32226; GB:M32227; GB:M32228; GB:M32229; GB:M32230; GB:M32231; GB:M32232; GB:M32233; GB:M32234; GB:M32235; GB:M32236; GB:M32237; GB:M32238; GB:M32239; GB:M32240; GB:M32241; GB:M32242; GB:M32243; GB:M32244; GB:M32245; GB:M32246; GB:M32247; GB:M32248; GB:M32249; GB:M32250; GB:M32251; GB:M32252; GB:M32253; GB:M32254; GB:M32255; GB:M32256; GB:M32257; GB:M32258; GB:M32259; GB:M32260; GB:M32261; GB:M32262; GB:M32263; GB:M32264; GB:M32265; GB:M32266; GB:M32267; GB:M32268; GB:M32269; GB:M32270; GB:M32271; GB:M32272; GB:M32273; GB:M32274; GB:M32275; GB:M32276; GB:M32277; GB:M32278; GB:M32279; GB:M32280; GB:M32281; GB:M32282; GB:M32283; GB:M32284; GB:M32285; GB:M32286; GB:M32287; GB:M32288; GB:M32289; GB:M32290; GB:M32291; GB:M32292; GB:M32293; GB:M32294; GB:M32295; GB:M32296; GB:M32297; GB:M32298; GB:M32299; GB:M32300; GB:M32301; GB:M32302; GB:M32303; GB:M32304; GB:M32305; GB:M32306; GB:M32307; GB:M32308; GB:M32309; GB:M32310; GB:M32311; GB:M32312; GB:M32313; GB:M32314; GB:M32315; GB:M32316; GB:M32317; GB:M32318; GB:M32319; GB:M32320; GB:M32321; GB:M32322; GB:M32323; GB:M32324; GB:M32325; GB:M32326; GB:M32327; GB:M32328; GB:M32329; GB:M32330; GB:M32331; GB:M32332; GB:M32333; GB:M32334; GB:M32335; GB:M32336; GB:M32337; GB:M32338; GB:M32339; GB:M32340; GB:M32341; GB:M32342; GB:M32343; GB:M32344; GB:M32345; GB:M32346; GB:M32347; GB:M32348; GB:M32349; GB:M32350; GB:M32351; GB:M32352; GB:M32353; GB:M32354; GB:M32355; GB:M32356; GB:M32357; GB:M32358; GB:M32359; GB:M32360; GB:M32361; GB:M32362; GB:M32363; GB:M32364; GB:M32365; GB:M32366; GB:M32367; GB:M32368; GB:M32369; GB:M32370; GB:M32371; GB:M32372; GB:M32373; GB:M32374; GB:M32375; GB:M32376; GB:M32377; GB:M32378; GB:M32379; GB:M32380; GB:M32381; GB:M32382; GB:M32383; GB:M32384; GB:M32385; GB:M32386; GB:M32387; GB:M32388; GB:M32389; GB:M32390; GB:M32391; GB:M32392; GB:M32393; GB:M32394; GB:M32395; GB:M32396; GB:M32397; GB:M32398; GB:M32399; GB:M32400; GB:M32401; GB:M32402; GB:M32403; GB:M32404; GB:M32405; GB:M32406; GB:M32407; GB:M32408; GB:M32409; GB:M32410; GB:M32411; GB:M32412; GB:M32413; GB:M32414; GB:M32415; GB:M32416; GB:M32417; GB:M32418; GB:M32419; GB:M32420; GB:M32421; GB:M32422; GB:M32423; GB:M32424; GB:M32425; GB:M32426; GB:M32427; GB:M32428; GB:M32429; GB:M32430; GB:M32431; GB:M32432; GB:M32433; GB:M32434; GB:M32435; GB:M32436; GB:M32437; GB:M32438; GB:M32439; GB:M32440; GB:M32441; GB:M32442; GB:M32443; GB:M32444; GB:M32445; GB:M32446; GB:M32447; GB:M32448; GB:M32449; GB:M32450; GB:M32451; GB:M32452; GB:M32453; GB:M32454; GB:M32455; GB:M32456; GB:M32457; GB:M32458; GB:M32459; GB:M32460; GB:M32461; GB:M32462; GB:M32463; GB:M32464; GB:M32465; GB:M32466; GB:M32467; GB:M32468; GB:M32469; GB:M32470; GB:M32471; GB:M32472; GB:M32473; GB:M32474; GB:M32475; GB:M32476; GB:M32477; GB:M32478; GB:M32479; GB:M32480; GB:M32481; GB:M32482; GB:M32483; GB:M32484; GB:M32485; GB:M32486; GB:M32487; GB:M32488; GB:M32489; GB:M32490; GB:M32491; GB:M32492; GB:M32493; GB:M32494; GB:M32495; GB:M32496; GB:M32497; GB:M32498; GB:M32499; GB:M32500; GB:M32501; GB:M32502; GB:M32503; GB:M32504; GB:M32505; GB:M32506; GB:M32507; GB:M32508; GB:M32509; GB:M32510; GB:M32511; GB:M32512; GB:M32513; GB:M32514; GB:M32515; GB:M32516; GB:M32517; GB:M32518; GB:M32519; GB:M32520; GB:M32521; GB:M32522; GB:M32523; GB:M32524; GB:M32525; GB:M32526; GB:M32527; GB:M32528; GB:M32529; GB:M32530; GB:M32531; GB:M32532; GB:M32533; GB:M32534; GB:M32535; GB:M32536; GB:M32537; GB:M32538; GB:M32539; GB:M32540; GB:M32541; GB:M32542; GB:M32543; GB:M32544; GB:M32545; GB:M32546; GB:M32547; GB:M32548; GB:M32549; GB:M32550; GB:M32551; GB:M32552; GB:M32553; GB:M32554; GB:M32555; GB:M32556; GB:M32557; GB:M32558; GB:M32559; GB:M32560; GB:M32561; GB:M32562; GB:M32563; GB:M32564; GB:M32565; GB:M32566; GB:M32567; GB:M32568; GB:M32569; GB:M32570; GB:M32571; GB:M32572; GB:M32573; GB:M32574; GB:M32575; GB:M32576; GB:M32577; GB:M32578; GB:M32579; GB:M32580; GB:M32581; GB:M32582; GB:M32583; GB:M32584; GB:M32585; GB:M32586; GB:M32587; GB:M32588; GB:M32589; GB:M32590; GB:M32591; GB:M32592; GB:M32593; GB:M32594; GB:M32595; GB:M32596; GB:M32597; GB:M32598; GB:M32599; GB:M32600; GB:M32601; GB:M32602; GB:M32603; GB:M32604; GB:M32605; GB:M32606; GB:M32607; GB:M32608; GB:M32609; GB:M32610; GB:M32611; GB:M32612; GB:M32613; GB:M32614; GB:M32615; GB:M32616; GB:M32617; GB:M32618; GB:M32619; GB:M32620; GB:M32621; GB:M32622; GB:M326

A;Residues: 752-831, 'PA', 834, 'E', 836-1005, 'K', 1007-1036, 'Q', 1038-1052, 'E', 1054-1068, 'T',
A;Cross-references: GB:100977; NID:G180812; PIDN:AAB3914.1; PID:G258774
A;Note: sequence extracted from NCBI backbone (NCBIP:117273); parts of this sequence were
A;Note: this translation is not annotated and this publication is not cited in GenBank
A;Note: mutant sequence associated with perinatal lethal hypochondrogenesis
E;Tiller, G.E.; Rmoin, D.L.; Murray, L.W.; Cohn, D.H.
Proc. Natl. Acad. Sci. U.S.A. 87, 3889-3893, 1990
A;Title: Tandem duplication within a type II collagen gene (COL2A1) exon in an individual
A;Reference number: S16502; MUID:90251662; PMID:2339128
A;Accession: S16502
A;Molecule type: DNA
A;Residues: 1164-1184, 'GPSKGAGANGPPI', 1185-1199, 'TIL2'
A;Cross-references: EMBL:M37126; NID:G180808; PIDN:AA52037.1; PID:G180809
A;Note: mutant sequence from a patient with spondyloepiphyseal dysplasia
R;Cheah, K.S.E.; Stoker, N.G.; Griffin, J.R.; Grosveld, F.G.; Solomon, E.
Proc. Natl. Acad. Sci. U.S.A. 82, 2555-2559, 1985
A;Title: Identification and characterization of the human type II collagen gene (COL2A1)
A;Reference number: A02858; MUID:85190534; PMID:3857598
A;Accession: A02858
A;Molecule type: DNA
A;Residues: 1032-1056, 'N', 1058-1068, 'T', 1070-1487, 'CHE'
A;Cross-references: GB:J00116; NID:G180395; PIDN:AAA51997.1; PID:G180396
R;Elima, K.; Vuorio, E.
Nucleic Acids Res. 15, 9499-9504, 1987
A;Title: Determination of the single polyadenylation site of the human pro-alpha-1(II) c
A;Reference number: A27280; MUID:88067771; PMID:2825137
A;Accession: A27280
A;Molecule type: mRNA
A;Residues: 1175-1487, 'ELI'
A;Cross-references: EMBL:X06268; NID:G30096; PIDN:CAA29604.1; PID:G30097
A;Experimental source: fetal epiphyseal cartilage
R;van der Rest, M.; Rosenberg, L.C.; Olsen, B.R.; Poole, A.R.
Biochem. J. 237, 923-925, 1986
A;Title: Chondrocalcin is identical with the C-propeptide of type II procollagen.
A;Reference number: A57033; MUID:87099927; PMID:3800925
A;Accession: A57033
A;Molecule type: protein
A;Residues: 'X5', 1244-1246, 'N', 1248, 'X', 1250-1265; 1295-1305; 1395-1408, 'VAN'
A;Note: chondrocalcin identified as released collagen 1(II) chain carboxyl-terminal prope
R;Strom, C.M.; Upholt, W.B.
Nucleic Acids Res. 12, 1025-1038, 1984
A;Title: Isolation and characterization of genomic clones corresponding to the human typ
A;Reference number: A21733; MUID:84118798; PMID:6320112
A;Accession: A21733
A;Molecule type: DNA
A;Residues: 1245-1295, 'STR1'
A;Cross-references: EMBL:X00339; EMBL:X00298; NID:G394699; PIDN:CAA25092.1; PID:G4378975
A;Accession: B21733
A;Molecule type: DNA
A;Residues: 894-909, 'PE', 'STR2'
A;Cross-references: GB:K01785; NID:G30035; PIDN:CAA25082.1; PID:G1335032
R;Nunez, A.M.; Francomano, C.; Young, M.F.; Martin, G.R.; Yamada, Y.
Biochemistry 24, 6343-6348, 1985
A;Title: Isolation and partial characterization of genomic clones coding for a human pro
A;Reference number: A24561; MUID:86104139; PMID:3002437
A;Accession: A24561
A;Molecule type: DNA
A;Residues: 1236-1358, 'NUN2'
A;Cross-references: GB:M12048; NID:G180017
A;Note: this translation is not annotated in GenBank entry HUMCCT2A, release 111.0
A;Note: the codons given for 1333-Lys (AGC) and 1350-Gly (GCA) are inconsistent with the
R;Sangiorgi, F.O.; Benson-Chanda, V.; de Wet, W.J.; Sobel, M.E.; Tsipouras, P.; Ramirez,
Nucleic Acids Res. 13, 2207-2222, 1985
A;Title: Isolation and partial characterization of the entire human pro alpha 1(II) coll
A;Reference number: I37249; MUID:85215609; PMID:2987845
A;Accession: S59491
A;Molecule type: DNA
A;Residues: 7-28, 'R', 99-114; 541-578; 786-802; 1055-1056, 'N', 1058-1068, 'T', 1070-1109; 1200-1
A;Accession: I8453
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 7-28, 'SAN2'

A;Cross-references: GB:M23759; NID:G180845; EMBL:X03320; GB:M24938; NID:G30104
A;Note: the GenBank PID is based on an incorrect reading frame
A;Accession: I37250
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 541-560, 'SAN2'
A;Cross-references: EMBL:X02378; GB:M23870; NID:G30107; PIDN:CAA26227.1; PID:G9229621
A;Accession: I37251

Query Match	71.7%	Score 2202;	DB 1;	Length 1487;
Best Local Similarity	71.9%	Pred. No. 5e-100;		
Matches 391;	Conservative 40;	Mismatches 113;	Indels 0;	Gaps 0;

QY 1 GSEGGVGRGPPGPPGAGAGPAGDPCADGCEPKAGKADGAPGIAGAPGPGFAGRPSGPE 60
Db 384 GPEGAGGPPGPPGPPGAGAGPAGDPCADGCEPKAGKADGAPGIAGAPGPGFAGRPSGPE 443
QY 61 GPGGPPGKGSPPGPPGAGAGPAGDPCADGCEPKAGKADGAPGIAGAPGPGFAGRPSGPE 120
Db 444 GATGPPGKGSPPGPPGAGAGPAGDPCADGCEPKAGKADGAPGIAGAPGPGFAGRPSGPE 503
QY 121 GPPGPPGKGSPPGPPGAGAGPAGDPCADGCEPKAGKADGAPGIAGAPGPGFAGRPSGPE 180
Db 504 GPPGPPGKGSPPGPPGAGAGPAGDPCADGCEPKAGKADGAPGIAGAPGPGFAGRPSGPE 563
QY 181 GSPGPPGKGSPPGPPGAGAGPAGDPCADGCEPKAGKADGAPGIAGAPGPGFAGRPSGPE 240
Db 564 GPPGPPGKGSPPGPPGAGAGPAGDPCADGCEPKAGKADGAPGIAGAPGPGFAGRPSGPE 623
QY 241 GPPGPPGKGSPPGPPGAGAGPAGDPCADGCEPKAGKADGAPGIAGAPGPGFAGRPSGPE 300
Db 624 GPPGPPGKGSPPGPPGAGAGPAGDPCADGCEPKAGKADGAPGIAGAPGPGFAGRPSGPE 683
QY 301 GPPGPPGKGSPPGPPGAGAGPAGDPCADGCEPKAGKADGAPGIAGAPGPGFAGRPSGPE 360
Db 684 GPPGPPGKGSPPGPPGAGAGPAGDPCADGCEPKAGKADGAPGIAGAPGPGFAGRPSGPE 743
QY 361 GPPGPPGKGSPPGPPGAGAGPAGDPCADGCEPKAGKADGAPGIAGAPGPGFAGRPSGPE 420
Db 744 GPPGPPGKGSPPGPPGAGAGPAGDPCADGCEPKAGKADGAPGIAGAPGPGFAGRPSGPE 803
QY 421 GPPGPPGKGSPPGPPGAGAGPAGDPCADGCEPKAGKADGAPGIAGAPGPGFAGRPSGPE 480
Db 804 GPPGPPGKGSPPGPPGAGAGPAGDPCADGCEPKAGKADGAPGIAGAPGPGFAGRPSGPE 863
QY 481 GPPGPPGKGSPPGPPGAGAGPAGDPCADGCEPKAGKADGAPGIAGAPGPGFAGRPSGPE 540
Db 864 GPPGPPGKGSPPGPPGAGAGPAGDPCADGCEPKAGKADGAPGIAGAPGPGFAGRPSGPE 923
QY 541 GKEG 544
Db 924 GKDG 927

RESULT 6

A41182
collagen alpha 1(II) chain precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 13-Aug-1999
C;Accession: A41182; A44885
R;Metaeasanta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.
J. Biol. Chem. 266, 16862-16869, 1991
A;Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and
A;Reference number: A41182; MUID:91358489; PMID:1885613
A;Accession: A41182
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-1419, 'MET'
A;Cross-references: GB:M65161
R;Cheah, K.S.; Lau, E.T.; Au, P.K.; Tam, P.P.
Development 111, 945-953, 1991
A;Title: Expression of the mouse alpha 1(II) collagen gene is not restricted to cartilage
A;Reference number: A44885; MUID:91347939; PMID:1879363

A>Note: the composition of peptides comprising residues 1-9 and 1-19 confirms the sequence
A>Note: this region (residues 651-671 above) probably corresponds to positions 1032-1052
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
ed and subsequently O-glycosylated.
C:Comment: The order of the nine CNBr peptides in the alpha 1(I) chain of rat skin colla
C:Comment: The complete chain contains 1052 residues.
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: blocked amino end; coiled coil; extracellular matrix; glycoprotein; hydroxyl
F:1/Modified site: blocked amino end (Glx) (probably pyrrolidone carboxylic acid) #status
F:9/Modified site: allysine (lys) #status experimental
F:103-424,547/Binding site: carbohydrate (lys) (covalent) #status experimental
F:103-424,547/Binding site: 5-hydroxylysine (lys) #status experimental
F:103/Modified site: 5-hydroxylysine (lys) (partial) #status experimental
F:424,547/Modified site: 5-hydroxylysine (lys) (partial) #status experimental

Query Match 70.2%; Score 2154; DB 1; Length 671;
Best Local Similarity 77.3%; Pred. No. 5.9e-98;
Matches 381; Conservative 28; Mismatches 36; Indels 48; Gaps 2;

QY 1 GSEGEVGRGEPGPGPAGAGPAGDPCADGEPGAKGADGAPGAGPFGFPGARGPSPE 60
|||
Db 200 GSEGPQVGRGEPGPGPAGAGPAGNPDQCPGAKGANGAPGAGPFGFPGARGPSQ 259
|||
QY 61 GPGGPPGPKGDSGEPCAPGSKGDTCAKCEPVPVGEPPGPGAGEGKPCARGEPPGTCLP 120
|||
Db 260 GFSGAPGPKNSGEFGAPGNKGDITAKCEPGPAGVQGPFGAGEGKCARGEFGPGLP 319
|||
QY 121 GPPGERGGPSGFFGADGVAGPKPAGERSGPGPAGKPSGPGAGPGEAGLPGAKGLT 180
|||
Db 320 GPPGERGGPSGFFGADGVAGPKPAGERSGPGPAGKPSGPGAGPGEAGLPGAKGLT 379
|||
QY 181 GSPGSPGDKTTPGCPAGEGDCRPPGPPPGARGEAGVMGPPGPKGAAGEPGKAGERGVP 240
|||
Db 380 GSPGSPGDKTTPGCPAGSBRGCPAGPPGARQAGVMGPPGPKGTAGPGEKAGERGVP 439
|||
QY 241 GPPGAVGAGKDXGAGAGPFGPPGAPAGERGEEGPGSPGFEGLPGPAGPGEAGKPGEE 300
|||
Db 440 GPPGAVGAGKDXGAGAGPFGPPGAPAGERGEEGPGSPGFEGLPGPAGPGEAGKPGZZ 499
|||
QY 301 GVPDGLGAPSGARGEPFGPERGVEGPPGAPGPGADGAPDGDGAKGDAGAPGPGSE 360
|||
Db 500 GVPDGLGAPSGARGEPFGPERGVEGPPGAPGPGADGAPDGDGAKGDAGAPGPGSQ 559
|||
QY 361 GAGPLEGMPGEGAAGLPGPKDRDAGPKGADSGPKDGVRLGTGTPGPPGAPAGDK 420
|||
Db 560 GAGPLZGNSGLZ-----GPPGSPGSGKBLGLPLGPIGPPGRGTGBAGPS 583
|||
QY 421 GSESGSAGTGTARGAPCDRGERGPPCPAGFAGPPGADGEPGAKGEPGAGKGDGPP 480
|||
Db 584 GFSGASGAGAPR-----GPPGASGPKBLGLPLGPIGPPGRGTGBAGPS 631
|||
QY 481 GPAGPAGPPGPIG 493
|||
Db 632 GPPGPGPPGPPG 644
|||

RESULT 9
B40333
collagen alpha 1(ii) chain precursor - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: B40333
R:Su, M.W.; Suzuki, H.R.; Bieker, J.J.; Solursh, M.; Ramirez, F.
J. Cell Biol. 115, 565-575, 1991
A:Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis em
A:Reference number: A40333; MUID:92011898; PMID:1918153
A:Accession: B40333
A:Status: preliminary
A:Molecule type: mRNA
A:Cross-references: GB:W63595
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F:137-96/Domain: von Willebrand factor type C repeat homology <WVC>

1 GSEPEGVRRGEPPPPGAGAAAGPAGPDGDEFGAKGADGAPGIAGAGPFGPAGRGPSGPE 60
389 GPEGAQGRGESGTPGSPGPGSAGSNPGTDGIFCAKGSSCASGTAGAPGFPGRGPGPQ 448
61 GPGGPPPKGDSCEPGAPGSKGDTGAKBEPGVGVGPPCPAGCEGKPGARGEPGPTGLP 120
449 GATPLGPKGETGDFGTAGFKGHEGPKGEIGSGAGQCAPGAPAGEEGKRGARGEPGAAGPL 508
121 GPPGERGPGSGRSGPFGADGVAGPKGPAGERGSPGAPGKSGPGEAGRPGEAGLPAGKGLT 180
509 GPPGERGAPGNRGFPFQDGLAGPKGAPGERGVPLGPGFKGNGCDPGRFPFGLFGARGLT 568
181 GSPGSPGDKTGPFGAGEDGRPGPPGPGARGEAGVWGVPPGKGAAGPFGKAGERGVP 240
569 GRPDAGPQGVKGFSGAAGEDGRFPFPQARGQCPGVWGFPGFKGANGFPKAGEKELG 628
241 GPPGAVGPAGKDXBAGAEAGPPGAPGAGERGEEGAPGPFEGPLPGPAGPPGARGKPEE 300
629 GAPLRLGPKDGTGTAQGNPGPAGPAGERGEQPPGSPGQGLPGPPGPGEGEGKPGDQ 688
301 GVRGDLGAPSGARGEGEPGGERGVSGPPGAPGACGADGAPGDDGAKGDAGAPCAPQSE 360
689 GVPGEAGAAGLVGPRGERGFPGERGSGSPQLQGRSLGPTGTGDKGATGSPGPNGAQ 748
361 GAPGLEGMFGBRGAAGLPFGPKGRDGDAGPKGADGSPKDGVRGLTGTPIGPPGAPAGPDK 420
749 GPPGLQGVFGERGAAGISGPKGDSGDTGKGPAGPKDQSGRLTGLPLGPPGAPGPKGEK 808
421 GESGSPGAPGTGARGAPDRGEGPPGCPAGFAGPPGADGEPGAKGEPGDAGAKGDAGPP 480
809 GESGSPGPGTIVGARGAPDRGEGNGPPGAPGAPGSDGQAGLKGQGESQKGDAGAP 868
481 GPAGPAGPFGTIDYVAPGAKGARGSAGPPGATGPPGAAGRVGPPGSGDAGPPGPPGA 540
869 GPGQSGGAPGPGQPTGVNGPKGARGAQGPAGATGPPGAAGRVGTPGPNNGNFPGPPGSA 928
541 GKEG 544
929 GKEG 932

RESULT 11
CGB01S
collagen alpha 1(I) chain - bovine (tentative sequence) (fragments) .
C:Species: Bos primigenius taurus (cattle)
C:Date: 24-Apr-1984 #sequence_revision 31-Dec-1993 #text_change 31-Mar-2000
C:Accession: A91193; A91229; A91387; A91388; A91389; A91390; A91391; A91392; A91393; A91394; A91395; A91396; A91397; A91398; A91399; A91400; A91401; A91402; A91403; A91404; A91405; A91406; A91407; A91408; A91409; A91410; A91411; A91412; A91413; A91414; A91415; A91416; A91417; A91418; A91419; A91420; A91421; A91422; A91423; A91424; A91425; A91426; A91427; A91428; A91429; A91430; A91431; A91432; A91433; A91434; A91435; A91436; A91437; A91438; A91439; A91440; A91441; A91442; A91443; A91444; A91445; A91446; A91447; A91448; A91449; A91450; A91451; A91452; A91453; A91454; A91455; A91456; A91457; A91458; A91459; A91460; A91461; A91462; A91463; A91464; A91465; A91466; A91467; A91468; A91469; A91470; A91471; A91472; A91473; A91474; A91475; A91476; A91477; A91478; A91479; A91480; A91481; A91482; A91483; A91484; A91485; A91486; A91487; A91488; A91489; A91490; A91491; A91492; A91493; A91494; A91495; A91496; A91497; A91498; A91499; A91500; A91501; A91502; A91503; A91504; A91505; A91506; A91507; A91508; A91509; A91510; A91511; A91512; A91513; A91514; A91515; A91516; A91517; A91518; A91519; A91520; A91521; A91522; A91523; A91524; A91525; A91526; A91527; A91528; A91529; A91530; A91531; A91532; A91533; A91534; A91535; A91536; A91537; A91538; A91539; A91540; A91541; A91542; A91543; A91544; A91545; A91546; A91547; A91548; A91549; A91550; A91551; A91552; A91553; A91554; A91555; A91556; A91557; A91558; A91559; A91560; A91561; A91562; A91563; A91564; A91565; A91566; A91567; A91568; A91569; A91570; A91571; A91572; A91573; A91574; A91575; A91576; A91577; A91578; A91579; A91580; A91581; A91582; A91583; A91584; A91585; A91586; A91587; A91588; A91589; A91590; A91591; A91592; A91593; A91594; A91595; A91596; A91597; A91598; A91599; A91600; A91601; A91602; A91603; A91604; A91605; A91606; A91607; A91608; A91609; A91610; A91611; A91612; A91613; A91614; A91615; A91616; A91617; A91618; A91619; A91620; A91621; A91622; A91623; A91624; A91625; A91626; A91627; A91628; A91629; A91630; A91631; A91632; A91633; A91634; A91635; A91636; A91637; A91638; A91639; A91640; A91641; A91642; A91643; A91644; A91645; A91646; A91647; A91648; A91649; A91650; A91651; A91652; A91653; A91654; A91655; A91656; A91657; A91658; A91659; A91660; A91661; A91662; A91663; A91664; A91665; A91666; A91667; A91668; A91669; A91670; A91671; A91672; A91673; A91674; A91675; A91676; A91677; A91678; A91679; A91680; A91681; A91682; A91683; A91684; A91685; A91686; A91687; A91688; A91689; A91690; A91691; A91692; A91693; A91694; A91695; A91696; A91697; A91698; A91699; A91700; A91701; A91702; A91703; A91704; A91705; A91706; A91707; A91708; A91709; A91710; A91711; A91712; A91713; A91714; A91715; A91716; A91717; A91718; A91719; A91720; A91721; A91722; A91723; A91724; A91725; A91726; A91727; A91728; A91729; A91730; A91731; A91732; A91733; A91734; A91735; A91736; A91737; A91738; A91739; A91740; A91741; A91742; A91743; A91744; A91745; A91746; A91747; A91748; A91749; A91750; A91751; A91752; A91753; A91754; A91755; A91756; A91757; A91758; A91759; A91760; A91761; A91762; A91763; A91764; A91765; A91766; A91767; A91768; A91769; A91770; A91771; A91772; A91773; A91774; A91775; A91776; A91777; A91778; A91779; A91780; A91781; A91782; A91783; A91784; A91785; A91786; A91787; A91788; A91789; A91790; A91791; A91792; A91793; A91794; A91795; A91796; A91797; A91798; A91799; A91800; A91801; A91802; A91803; A91804; A91805; A91806; A91807; A91808; A91809; A91810; A91811; A91812; A91813; A91814; A91815; A91816; A91817; A91818; A91819; A91820; A91821; A91822; A91823; A91824; A91825; A91826; A91827; A91828; A91829; A91830; A91831; A91832; A91833; A91834; A91835; A91836; A91837; A91838; A91839; A91840; A91841; A91842

R;Fietzek, P.P.; Rexrodt, F.W.; Hopper, K.E.; Kuehn, K.
 Eur. J. Biochem. 38, 396-400, 1973
 A:Title: The covalent structure of collagen. 2. The amino-acid sequence of alpha1-CB7 fi
 A:Reference number: A91211; MUID:74086118; PMID:4359390
 A:Accession: A91211
 A:Molecule type: protein
 A:Residues: 255-562 <FI3>
 A:Experimental source: skin
 R;Wendt, P.; Mark, K.V.D.; Rexrodt, F.; Kuehn, K.
 Eur. J. Biochem. 30, 169-183, 1972
 A:Title: The covalent structure of collagen. The amino-acid sequence of the 112 residues
 A:Reference number: A91201; MUID:73042276; PMID:4343808
 A:Accession: A91201
 A:Molecule type: protein
 A:Residues: 563-675 <WEN>
 A:Experimental source: skin
 R;Fietzek, P.P.; Rexrodt, F.W.; Wendt, P.; Stark, M.; Kuehn, K.
 Eur. J. Biochem. 30, 163-168, 1972
 A:Title: The covalent structure of collagen. Amino acid sequence of peptide alpha1-CB6-C
 A:Reference number: A91200; MUID:73042275; PMID:4343807
 A:Accession: A91200
 A:Molecule type: protein
 A:Residues: 676-758 <FI4>
 A:Experimental source: skin
 A:Note: Pro-726 is the only 3-hydroxyproline and the only hydroxylated proline in positio
 R;Rautenberg, J.; Fietzek, P.; Rexrodt, F.; Becker, U.; Stark, M.; Kuehn, K.
 FEBS Lett. 21, 75-79, 1972
 A:Title: The amino acid sequence of the carboxyterminal nonhelical cross link region of
 A:Reference number: A43048
 A:Accession: A43048
 A:Molecule type: protein
 A:Residues: 759-779 <RA2>
 A:Experimental source: skin
 C:Comment: Lysines at positions 115, 124, 274, 346, 424, 496, 658, and 670 may be hydrox
 C:Comment: Prolines in the third position of the tripeptide repeating unit (G-X-Y) are h
 C:Comment: The order of the eight CNBr peptides in the alpha 1(I) chain of bovine skin c
 9, 149, 268, and 217 residues.
 C:Comment: The complete chain contains 1052 residues.
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 C:Keywords: coiled coil; extracellular matrix; glycoprotein; pyroglutamic acid; trimer;
 F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
 Query Match 67.6%; Score 2076.5; DB 1; Length 779;
 Best Local Similarity 78.0%; Pred. No. 3.7e-94;
 Matches 382; Conservative 18; Mismatches 51; Indels 39; Gaps 6;
 QY 55 GPSGEGGPGPPKSGSGPGAGSGKDTGAKGPGVGVGPPGPGAGEGKPGARGGP 114
 Db 17 GFMGPGPRGLFPG-----PGA-----PPQGFQGGP---GEPGEPGASGPM 55
 QY 115 GPTGLPFPGERGPGSRGPPGADGVAGPKGPPAGERGSGPPGAPKSGPGEAGRPAGLP 174
 Db 56 GPRGPPGPKNGD-----DGEAGKGRGPRGPGPPQCAR-----GLPGTAGLP 100
 QY 175 GAKLTGSPGPGDGTGTGPPGAGEDGRPGPPGPPGARGAGVWGPFGPKGAGEPGKA 234
 Db 101 GMKGHRGSLDGAAGDAGPAGPKGEPGSGENGAPGQMGPRGLPGFPFGKGAAGPFGKA 160
 QY 235 GERGVPPPGAVGAGKDGAGAGGPPGPPGAPAGERGEGPAGSGPSGFEGLPGPAGPGEA 294
 Db 161 GERGVPPPGAVGAGKDGAGAGGPPGPPGAPAGERGEGPAGSGPSGFEGLPGPAGPGEA 220
 QY 295 GXPGEGVPGDLGAPGSGARGPGFPFGERGVEGPPGAPGADGAPGDGAKGDAGAP 354
 Db 221 GKPGEQVPGDLGAPGSGARGPGFPFGERGVEGPPGAPGAGPAGPAGPAGPAGPAGPAG 280
 QY 355 GAFSGEAGLEGMPGPRGAGAGPGPKGDRGAGPKGDSCKDVRLGTPIGPFGPA 414
 Db 281 GAFSGQAPLQGMFGERGAAGPGPKGDRGAGPKGDSCKDVRLGTPIGPFGPA 340
 QY 415 GAPDKGEGSPGAPGTGARGAGDGRGPPGPPGAPGAGPAGPAGPAGPAGPAGPAG 474
 Db 341 GAPDKGEGSPGPA---GTRGAPGRGPRGPPGAPGAGPAGPAGPAGPAGPAGPAG 397

QY 475 GDAGPPGAPGAPGPPGPIGDVCGAPGAKGARGAGPGATGFGAAGRGVGGPPGSDAGPP 534
 Db 398 GDAGPPGAPGAPGPPGPIGCVGAPGKARGAGPGATGFGAAGRGVGGPPGSDAGPP 457
 QY 535 GPFPGAPGKEG 544
 Db 458 GPFPGAPGKEG 467
 RESULT 12
 CGHU2V
 collagen alpha 2(V) chain precursor - human
 C:Species: Homo sapiens (man)
 C:Date: 31-Jul-1989 #sequence revision 28-Jul-1995 #text change 31-Dec-2000
 C:Accession: A31427; A54555; A54643; A25874; I55239; I59025; A25374; A30017
 R;Woodbury, D.; Benson-Chanda, V.; Ramirez, F.
 J. Biol. Chem. 264, 2735-2738, 1989
 A:Title: Amino-terminal propeptide of human pro-alpha2(V) collagen conforms to the struct
 A:Reference number: A31427; MUID:89123368; PMID:2914927
 A:Accession: A31427
 A:Molecule type: mRNA
 A:Residues: 1-463 <NOO>
 A:Cross-references: GB:J04478; NID:gl79697; PIDN:AAA51859.1; PID:gl79698
 A:Experimental source: placenta
 R;Greenspan, D.S.; Lee, S.T.; Lee, B.S.; Hoffman, G.G.
 Gene Expr. 1, 29-39, 1991
 A:Title: Homology between alpha2(V) and alpha1(III) collagen promoters and evidence for i
 A:Reference number: A54555; MUID:92314691; PMID:1820205
 A:Accession: A54555
 A:Molecule type: DNA
 A:Residues: 1-32 <GRE>
 A:Cross-references: GB:M58525; NID:gl80834; PIDN:AAC41699.1; PID:G553235
 R;Moradi-Ameli, M.; Rousseau, J.C.; Kleman, J.P.; Champliand, M.F.; Boutillon, M.M.; Ber
 Eur. J. Biochem. 231, 987-995, 1994
 A:Title: Diversity in the processing events at the N-terminus of type-V collagen.
 A:Reference number: S43642; MUID:94237164; PMID:8181482
 A:Accession: S43643
 A:Molecule type: protein
 A:Residues: 288-291, 'P', 293-294, 'X', 296-297, 606, 'X', 608-617 <MOR>
 R;Weil, D.; Bernard, M.; Gargano, S.; Ramirez, F.
 Nucleic Acids Res. 15, 181-198, 1987
 A:Title: The pro alpha 2(V) collagen gene is evolutionarily related to the major fibrill
 A:Reference number: A25874; MUID:87146331; PMID:3029669
 A:Accession: A25874
 A:Molecule type: mRNA; DNA
 A:Residues: 398-1496 <WEI>
 A:Cross-references: GB:X04758; NID:g29588; PIDN:CAA28454.1; PID:gl340175
 A:Experimental source: rhabdomyosarcoma cell line
 R;Myers, J.C.; Loidl, H.R.; Stolle, C.A.; Seyer, J.M.
 J. Biol. Chem. 260, 5533-5541, 1985
 A:Title: Partial covalent structure of the human alpha 2 type V collagen chain.
 A:Reference number: I55239; MUID:85182703; PMID:2985598
 A:Accession: I55239
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1002-1226 <BE2>
 A:Cross-references: GB:M10956; NID:gl80427; PIDN:AAA52007.1; PID:gl80428
 R;Emanuel, B.S.; Cammizaro, L.A.; Seyer, J.M.; Myers, J.C.
 Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985
 A:Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm
 A:Reference number: I59025; MUID:85216505; PMID:3858826
 A:Accession: I59025
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1003-1034 <RES>
 A:Cross-references: GB:M11135; NID:gl79693; PIDN:AAA51857.1; PID:gl79694
 A:Note: part of this sequence were determined by protein sequencing
 R;Myers, J.C.; Loidl, H.R.; Seyer, J.M.; Dion, A.S.
 J. Biol. Chem. 260, 11216-11222, 1985
 A:Title: Complete primary structure of the human alpha-2 type V procollagen COOH-terminal
 A:Reference number: A25374; MUID:85289337; PMID:2411731

QY	121	GPPEGGCGSRGGPGGADGAGKGPAGRGSGPCGAGPKGSGPCEAGRPGCEAGLPGAGGLT	180
DB	513	GVGVGERGAPGNRGPPGSDGLPGPKGAOGERGPGVSSGPKSGQDQGRPGEPGLPGAGLGT	572
QY	181	GSFGSGFGDGTGPPGAGEDGRPPGPPGARGAEGVMGFPGPKGAAGEPGKAGAGVGP	240
DB	573	GNFGVQGGEGKGLPLGAPGEDGRDPPGSGTIGKQPGTMTGLPGPKGSGNDGPKPGGAGNP	632
QY	241	GPAGVAGPAGKDXGAGAGPPGAPGAGRGEGSGPAGSGPGEGLPGAPGPPGCEAGKPGEE	300
DB	633	GVFGQRGAPGKDXGKGVPGPPGGLRGERGEGQPGTGFQGHPRGPPPPGEGKPGDQ	692
QY	301	GVPGDLGAPGSGARGSPGPPGGERGVGPPGPPAGPPGADGAPGDDGKAGDAGAPAGSGE	360
DB	693	GVPGGPGAVGGLGPRGERGNPGERGPCTITGLPCEKMGAGHGPGDFGKSGPFGTFGDT	752
QY	361	GAPGLEGWGEGRGAGLPGPKGRDGDAGPKGADGSPKSGKVRGLTGPIGPPGAPGAPGDK	420
DB	753	GPPLQLGNGWGGRTAGTTPGPKDRGGTIGEKAGSTAGNDAGGLPGPLGPPGAPGLLGEK	812
QY	421	GSFGSPGAGTQARGAPGDRGEPGPPGPPAGFPAGPGADGEPGAKGEPGDGAKGDAGPP	480
DB	813	GPFGPRGLVGPSPGRGNPGSRGNGTGAVFAGPQSGDQGPVKGEPGPGKGDAGSP	872
QY	481	GPAGPAGPPGPIGDVGAPGAKGARGSGAPGATGFFGCAAGRVGPPGPGSDAGPDPGPGA	540
DB	873	GFOGLAGSPGPHGPNVGPGLKGRGTQPPGATGFFGSAAGRVGPPGAPGAPGAPGLGPP	932
QY	541	GKEG 544	
DB	933	GKEG 936	

RESULT 13

A43291

collagen alpha 2(I) chain precursor - mouse

C:Species: Mus musculus (house mouse)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: A43291; A54328

R:Phillips, C.L.; Morgan, A.L.; Lever, L.W.; Wenstrup, R.J.

Genomics 13, 1345-1346, 1992

A:Title: Sequence analysis of a full-length cDNA for the murine pro alpha 2(I)

A:Reference number: A43291; MUID:92372043; PMID:1505972

A:Accession: A43291

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-1373 <PH1>

A:Cross-references: NID:G50488; PIDN:CAA41205.1; PID:G50489

A>Note: sequence extracted from NCBI backbone (NCBIP:112027)

R:Phillips, C.L.; Lever, L.W.; Pinnell, S.R.; Quarles, L.D.; Wenstrup, R.J.

J. Invest. Dermatol. 97, 980-984, 1991

A:Title: Construction of a full-length murine Proalpha2(I) collagen cDNA by th

A:Reference number: A54328; MUID:92084969; PMID:1748823

A:Accession: A54328

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-110 <PH2>

C:Genetics:

A:Gene: COL1A2

C:Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal

F;1145-1373/Domain: fibrillar collagen carboxy-terminal homology <FCC>

	Query Match	62.4%	Score 1915,	DB 1:	Length 1373;
	Best Local Similarity	64.2%;	Pred.No.	3.7e-86;	
	Matches 349;	Conservative 45;	Mismatches 150;	Indels 0;	Gaps 0;
Qy	1 GSEPFVGRGEPGGPAGAAAGDPADPGACDCEPGAAGKAGADCAPGTAGAPFGCARGSPGE	60			
Dd	280 GPAGFAPRGEVGVLGLSGVPVGPNGNPTGNLTAKAGATLFGVAGAPGLEPPRGIPGA	339			
Qy	61 GPGFPPFKDGSDSFPAGSGKDGTGARGFPGPVAGEPPGAPSEGKPCARGEPPTGLP	120			

```

Db      340 GAAGATGAGLVGEPGAGKESGNKGPESVGAQGPFGSGBEGKKGSGPGEAGSAGPA 399
QY      121 GPPGRRGGSGRGFFGADGVAQPKGAPGERSGPGKSGFAGRGPRGEGAGLPCAAGLT 180
Db      400 GPPGLRGSPGSLGFCADGACGAGVWGPFGNRSSTGAGIRGPNAGDAGRCPEGLMGRPLG 459
QY      181 GSPGSGPDGKTPGPPGAGEDGRCPPGPPGARGGAGVWGPFGKGAERGKAGRCVP 240
Db      460 GSGVNGVSGKEGPGVGLPCIDRRPPTGAPRGEGAGNIGFPFGKSGDGPCKGERGHP 519
QY      241 GPPGAVGAGKDEAGAGGPPGAPGAGRGEEGEPAGSPGFEGLFGGAGPPGEGEACKPCEE 300
Db      520 GLAGARGAGPDGNNAGAQGPFGQVQCKGEGQGGAGFPFGQLPGPSGTTGEVKGPER 579
QY      301 GVPGLGARGSGARGGFTGGERVEGPPGAPGPPGADGAPDDGAKGDAGATGATGSE 360
Db      580 GLPGEFLGPGAPRGERTGPGESGAAGPSGPIGSRGPGSAGPDGNNKGEAGAVGAPGSA 639
QY      361 GAPGLEGMPPGERRGAAGLPGPKGDRDAGPKGADGSPGKDGVRGLTGTGTPGPPGAPGADGK 420
Db      640 GASGPGGLPGERGAAGIPGKGEKGETGLRGDTGNTGFDGARGIPGAVGAPGAPGAGS 699
QY      421 GESGSPGAPGTGARGAGDGRGEPGPPGAPGADGAPGAGDGPAGKAGDAGADGAP 480
Db      700 GEAGAAGSPGAPRGSPGGERGEGVGPAGPAGPAGAGAGQFAGKAGKGTGKPGKENGIV 759
QY      481 GPAGPAGPPGPIGDVGAQKAGARGSGAGPGATGFPGAAGRVGPPGSGDAGPPGPGPA 540
Db      760 GPTGSGVGAAGSGPNGPFGPVGSRDGGPGWGTGFPGAAGRTGPPGSGIAGPPGPGAA 819
QY      541 GKEG 544
Db      820 GKEG 823

RESULT 14
I49607
procollagen type V alpha 2 - mouse
C:Species: Mus musculus (house mouse)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 13-Aug-1999
C:Accession: I49607
R:Andrikopoulos, K.; Suzuki, H.R.; Solursh, M.; Ramirez, F.
Dev. Dyn. 195, 113-120, 1992
A:Title: Localization of pro-alpha 2(V) collagen transcripts in the tissues of the devel
A:Reference number: I49607; MUID:93214071; PMID:1297453
A:Accession: I49607
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1497 <RES>
A:CROSS-references: GB:L02918; NID:G309180; PIDN:AAA37440.1; PID:G309181
C:Genetics:
A:Gene: Col5a-2
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
F:39-98/Domain: von Willebrand factor type C repeat homology <VMC>
F:1270-1497/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match      62.2%; Score 1910; DB 2; Length 1497;
Best Local Similarity 63.8%; Pred. No. 6.9e-86;
Matches 347; Conservative 39; Mismatches 158; Indels 0; Gaps 0;

QY      1 GSEGEVGRGPPGPPGAGAGPAGDGPAGDEPGKAGDAGPAGPAGPAGPAGPAGPAGPAGP 60
Db      394 GPEGPGQGRGTGPPGAGSGLGAVGTGTFGRKCATSGATSGTSPGGLAGPFGPFGPQ 453
QY      61 GPGGPPGKSGSPGAPSGKDTGAKGEGPCPVGVEGPPGAPGAGGKPGARGSPGTPGLP 120
Db      454 GSTPGQIRGSGDPPGVFGFGEAGPAGKGPFGHGIQPIGPPGEEGKRGPRGPGTGVPP 513
QY      121 GPPGERGGSGRGFPAGDGVAGPAGRGSPGAPGKSGPAGPAGPAGPAGPAGPAGPAGPAG 180
Db      514 GPMGERAGPNRGFPDGLGPKAGAGRGKPVGSSGKGGQDGPAGPAGPAGPAGPAGPAG 573
QY      181 GSPGSGPDGKTPGPPGAGEDGRCPPGPPGARGGAGVWGPFGKGAERGKAGRCVP 240

```

```

Db      574 GNPVGQPGKGLGPGAGPAGEDGRPPGPGSIGTRGPGSGMGVFPKSGSDGLGKPGEAGNA 633
QY      241 GPPGAVGAPGKDEAGAGGPPGAPGARGGEGPAGSPGFEGLFGGAGPPGEGEACKPCEE 300
Db      634 GVPGQKAGKDEGVEGSPVFPVPGGLAGRGAGPFGTGTGFGQLGPPGPPGEGGKAGDQ 693
QY      301 GVPGLGARGSGARGGFTGGERVEGPPGAPGPPGADGAPDDGAKGDAGADGAPGAPGSE 360
Db      694 GVPGEFGVGLPGRGNGPGERGEGITGLPGKGVAGGHPDGPKNPGTGTIGDT 753
QY      361 GAPGLEGMPPGERRGAAGLPGPKGDRDAGPKGADGSPGKDGVRGLTGTGTPGPPGAPGADGK 420
Db      754 GPPGLQGMFGERGIAGTGPFPKDRGIGEKGAEGTAGNDGARGLPGPLGPPGAGLLGAP 813
QY      421 GESGSPGAPGTGARGAGDGRGEPGPPGAPGADGAPGAGDGPAGKAGDAGADGAP 480
Db      814 GPPGPRGLVFPGRSGNPGSRGNGTGAAGVAGPAGQGGQPGVKGEGEPGQKGDAGSP 873
QY      481 GPAGPAGPPGPIGDVGAQKAGARGSGAGPPGATGFPGAAGRVGPPGSGDAGPPGPGPA 540
Db      874 GPGLAGSPGPPHGVPLGKGRGTGQPPGATGFPGAGRVGPPGAPGAPGAPGAPG 933
QY      541 GKEG 544
Db      934 GKEG 937

RESULT 15
S59856
collagen alpha 1(III) chain precursor - mouse
C:Species: Mus musculus (house mouse)
C>Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 13-Aug-1999
C:Accession: S59856; S62120; S16373
R:Toman, P.D.; de Crombrughe, B.
Gene 147, 161-168, 1994
A:Title: The mouse type-III procollagen-encoding gene: genomic cloning and complete DNA s
A:Reference number: S59856; MUID:95011609; PMID:7926795
A:Accession: S59856
A:Molecule type: DNA
A:Residues: 1-1464 <TOM>
A:CROSS-references: EMBL:X52046
R:Toman, D.
submitted to the EMBL Data Library, November 1994
A:Reference number: S62120
A:Accession: S62120
A:Molecule type: DNA
A:Residues: 1-866, G', 868-1464 <TOA>
A:CROSS-references: EMBL:X52046; NID:G575321; PIDN:CAA36279.1; PID:G575322
R:Metzgerant, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.
Biochim. Biophys. Acta 1089, 241-243, 1991
A:Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.
A:Reference number: S16176; MUID:91274355; PMID:2054384
A:Accession: S16373
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1442-1464 <MET>
A:CROSS-references: EMBL:X57983; NID:G50476; PIDN:CAA41048.1; PID:G50477
C:Genetics:
A:Insertions: 29/1, 95/3; 112/3; 150/3; 175/3; 193/3; 211/3; 229/3; 247/3; 265/3; 283/3; 298/3; 673/3; 706/3; 742/3; 760/3; 778/3; 796/3; 814/3; 850/3; 868/3; 886/3; 940/3; 976/3;
C:Superfamily: collagen alpha 1(II) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: coiled coil; extracellular matrix
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-154/Domain: propeptide #status predicted <PRO>
F:32-92/Domain: von Willebrand factor type C repeat homology <VMC>
F:155-1464/Product: collagen alpha 1(III) chain #status predicted <MAT>
F:1236-1464/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match      62.1%; Score 1908; DB 2; Length 1464;
Best Local Similarity 63.3%; Pred. No. 8.5e-86;
Matches 346; Conservative 37; Mismatches 158; Indels 6; Gaps 2;

```

Search completed: September 24, 2004, 11:11:47
Job time : 22.9509 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2004, 11:06:55 ; Search time 13.7264 Seconds
(without alignments)
2063.622 Million cell updates/sec

Title: US-10-658-989A-4
Perfect score: 3070
Sequence: 1 GSEGPGEVGRGPGPPGAGA.....PGPSGDAGPRGPGAGKEG 544

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2966	96.6	1464	1 CA11_HUMAN	P02452 homo sapien
2	2940	95.8	1460	1 CA11_CANFA	O9x5j7 canis famil
3	2857	93.1	1453	1 CA11_MOUSE	P11087 mus musculu
4	2775	90.4	1453	1 CA11_CHICK	P02457 gallus gall
5	2202	71.7	1418	1 CA12_HUMAN	P02458 homo sapien
6	2192	71.4	1459	1 CA12_MOUSE	P28481 mus musculu
7	2151	70.1	671	1 CA11_RAT	P02454 rattus norv
8	2076.5	67.6	779	1 CA11_BOVIN	P02453 bos taurus
9	1949	63.5	1364	1 CA21_BOVIN	P02465 bos taurus
10	1934.5	63.0	1262	1 CA13_CHICK	P21205 gallus gall
11	1834	63.0	1372	1 CA21_RAT	P02466 rattus norv
12	1921.5	62.6	1496	1 CA25_HUMAN	P05997 homo sapien
13	1915	62.4	1372	1 CA21_MOUSE	Q01149 mus musculu
14	1905.5	62.1	1049	1 CA13_BOVIN	P04258 bos taurus
15	1901	61.9	1466	1 CA13_HUMAN	P02461 homo sapien
16	1899	61.9	1464	1 CA13_MOUSE	P08121 mus musculu
17	1898	61.8	1366	1 CA21_CANFA	O46392 canis famil
18	1880	61.2	1366	1 CA21_HUMAN	P08123 homo sapien
19	1834.5	59.8	1355	1 CA21_RANCA	O42350 rana catesb
20	1822	59.3	1362	1 CA21_CHICK	P02467 gallus gall
21	1789	58.3	1356	1 CA21_ONCMY	O93484 oncorhynch
22	1496	48.7	747	1 CA12_BOVIN	P02459 bos taurus
23	1454.5	47.4	1838	1 CA15_HUMAN	P20908 homo sapien
24	1449	47.2	1806	1 CA1B_HUMAN	P12107 homo sapien
25	1447	47.1	1745	1 CA1B_HUMAN	P25940 homo sapien
26	1432	46.6	1804	1 CA1B_MOUSE	Q61245 mus musculu
27	1423.5	46.4	1736	1 CA2B_HUMAN	P13942 homo sapien
28	1414.5	46.1	1736	1 CA2B_MOUSE	Q64739 mus musculu
29	1410	45.9	1027	1 CAFF_RIPPA	P10754 riftia pach
30	1387	45.2	1685	1 CA5A_HUMAN	P29400 homo sapien
31	1375.5	44.8	2944	1 CA17_HUMAN	Q02388 homo sapien
32	1372	44.7	1763	1 CA24_ASCSU	P27393 ascaris suu
33	1357.5	44.2	1690	1 CA44_HUMAN	P53420 homo sapien

ALIGNMENTS

RESULT 1

CALL_HUMAN	CALL_HUMAN	STANDARD;	PRT;	1464 AA.
AC P02452; Q14037; Q15176;	AC	P02452; Q14037; Q15176;		
DT 21-JUL-1986 (Rel. 01, Created)	DT	21-JUL-1986 (Rel. 01, Created)		
DT 01-MAR-1989 (Rel. 10, Last sequence update)	DT	01-MAR-1989 (Rel. 10, Last sequence update)		
DT 10-OCT-2003 (Rel. 42, Last annotation update)	DT	10-OCT-2003 (Rel. 42, Last annotation update)		
DE Collagen alpha 1(I) chain precursor.	DE	Collagen alpha 1(I) chain precursor.		
GN COL1A1.	GN	COL1A1.		
OS Homo sapiens (Human).	OS	Homo sapiens (Human).		
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX NCBI_TaxID=9606;	OX	NCBI_TaxID=9606;		
RN [1]	RN	[1]		
RP SEQUENCE OF 1-472 FROM N.A.	RP	SEQUENCE OF 1-472 FROM N.A.		
RX MEDLINE=89025644; PubMed=3178743;	RX	MEDLINE=89025644; PubMed=3178743;		
RA Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,	RA	Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,		
RA Jaenisch R., Prockop D.J.;	RA	Jaenisch R., Prockop D.J.;		
RT "Structure of a full-length cDNA clone for the prepro alpha 1(I)	RT	"Structure of a full-length cDNA clone for the prepro alpha 1(I)		
RT chain of human type I procollagen.";	RT	chain of human type I procollagen.";		
RL Biochem. J. 253:919-922(1988).	RL	Biochem. J. 253:919-922(1988).		
RN [2]	RN	[2]		
RP SEQUENCE OF 1-181 FROM N.A.	RP	SEQUENCE OF 1-181 FROM N.A.		
RX MEDLINE=84270697; PubMed=6462220;	RX	MEDLINE=84270697; PubMed=6462220;		
RA Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morabito M.,	RA	Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morabito M.,		
RA Myers J., Williams C., Ramirez F.;	RA	Myers J., Williams C., Ramirez F.;		
RT "Human pro alpha 1(I) collagen gene structure reveals evolutionary	RT	"Human pro alpha 1(I) collagen gene structure reveals evolutionary		
RT conservation of a pattern of introns and exons.";	RT	conservation of a pattern of introns and exons.";		
RL Nature 310:337-340(1984).	RL	Nature 310:337-340(1984).		
RN [3]	RN	[3]		
RP SEQUENCE OF 162-301.	RP	SEQUENCE OF 162-301.		
RC TISSUE=Skin;	RC	TISSUE=Skin;		
RX MEDLINE=71038625; PubMed=5529814;	RX	MEDLINE=71038625; PubMed=5529814;		
RA Click E.N., Bornstein P.;	RA	Click E.N., Bornstein P.;		
RT "Isolation and characterization of the cyanogen bromide peptides from	RT	"Isolation and characterization of the cyanogen bromide peptides from		
RT the alpha 1 and alpha 2 chains of human skin collagen.";	RT	the alpha 1 and alpha 2 chains of human skin collagen.";		
RL Biochemistry 9:4699-4706(1970).	RL	Biochemistry 9:4699-4706(1970).		
RN [4]	RN	[4]		
RP SEQUENCE OF 263-268.	RP	SEQUENCE OF 263-268.		
RC TISSUE=Skin;	RC	TISSUE=Skin;		
RX MEDLINE=71001508; PubMed=4319110;	RX	MEDLINE=71001508; PubMed=4319110;		
RA Morgan P.H., Jacobs H.G., Segrest J.P., Cunningham L.W.;	RA	Morgan P.H., Jacobs H.G., Segrest J.P., Cunningham L.W.;		
RT "A comparative study of glycopeptides derived from selected	RT	"A comparative study of glycopeptides derived from selected		
RT vertebrate collagens. A possible role of the carbohydrate in fibril	RT	vertebrate collagens. A possible role of the carbohydrate in fibril		
RT formation.";	RT	formation.";		
RL J. Biol. Chem. 245:5042-5048(1970).	RL	J. Biol. Chem. 245:5042-5048(1970).		
RN [5]	RN	[5]		
RP SEQUENCE OF 425-1464 FROM N.A.	RP	SEQUENCE OF 425-1464 FROM N.A.		
RX MEDLINE=84080385; PubMed=6689127;	RX	MEDLINE=84080385; PubMed=6689127;		
RA Bernard M.P., Chu M.-L., Myers J.C., Ramirez F., Eikenberry E.F.,	RA	Bernard M.P., Chu M.-L., Myers J.C., Ramirez F., Eikenberry E.F.,		
RA Prockop D.J.;	RA	Prockop D.J.;		
RT "Nucleotide sequences of complementary deoxyribonucleic acids for the	RT	"Nucleotide sequences of complementary deoxyribonucleic acids for the		
RT pro alpha 1 chain of human type I procollagen. Statistical evaluation	RT	pro alpha 1 chain of human type I procollagen. Statistical evaluation		
RT of structures that are conserved during evolution.";	RT	of structures that are conserved during evolution.";		
RL Biochemistry 22:5213-5223(1983).	RL	Biochemistry 22:5213-5223(1983).		
RN [6]	RN	[6]		

Q28083 bos taurus
Q01955 homo sapien
Q07092 homo sapien
P02462 homo sapien
P17140 caenorhabdi
P02463 mus musculu
P08572 homo sapien
P39060 homo sapien
Q28447 canis famil
Q14050 homo sapien
Q14993 homo sapien
Q05722 mus musculu

34 1356 44.2 911 1 CA1B_BOVIN
35 1344.5 43.8 1670 1 CA14_HUMAN
36 1339.5 43.6 1603 1 CA1F_HUMAN
37 1320 43.0 1669 1 CA14_HUMAN
38 1310.5 42.7 1758 1 CA24_CAEEL
39 1310 42.7 1669 1 CA14_MOUSE
40 1286.5 41.9 1712 1 CA24_HUMAN
41 1281.5 41.7 1516 1 CA1H_HUMAN
42 1279.5 41.7 754 1 CA5H_CANFA
43 1277 41.6 684 1 CA39_HUMAN
44 1276.5 41.6 1143 1 CA1I_HUMAN
45 1276 41.6 921 1 CA19_MOUSE

RP SEQUENCE OF 1229-1454 FROM N.A.
RC TISSUE-Bone;
RX MEDLINE=89124208; PubMed=3340531;
RA Maekelae J.K., Raasina M., Virta A., Vuorio E.;
RT "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide
RT domain.";
RL Nucleic Acids Res. 16:349-349(1988).
RN [7]
RP SEQUENCE OF 1-34 FROM N.A.
RX MEDLINE=89097389; PubMed=3480516;
RA Bornstein J., McKay J., Morishima J.K., Devarayalu S., Gelinas R.E.;
RT "Regulatory elements in the first intron contribute to
RT transcriptional control of the human alpha 1(I) collagen gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:8869-8873(1987).
RN [8]
RP SEQUENCE OF 1-44 FROM N.A.
RX MEDLINE=89130970; PubMed=2857713;
RA Chu M.-L., de Wet W.J., Bernard M.P., Ramirez F.;
RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
RT Promoter structure, Alu repeats, and polymorphic transcripts.";
RL J. Biol. Chem. 260:2315-2320(1985).
RN [9]
RP SEQUENCE OF 1-44 FROM N.A.
RX MEDLINE=89033098; PubMed=2822714;
RA Rossouw C.M.S., Vergeer W.P., du Plooy S.J., Bernard M.P., Ramirez F.,
RA de Wet W.J.;
RT "DNA sequences in the first intron of the human pro-alpha 1(I)
RT collagen gene enhance transcription.";
RL J. Biol. Chem. 262:15151-15157(1987).
RN [10]
RP REVIEW ON VARIANTS.
RX MEDLINE=91184577; PubMed=2010058;
RA Kuvaliemi H., Tromp G., Prockop D.J.;
RT "Mutations in collagen genes: causes of rare and some common diseases
RT in humans.";
RL PASEB J. 5:2052-2060(1991).
RN [11]
RP REVIEW ON VARIANTS.
RX MEDLINE=97255959; PubMed=9101290;
RA Kuvaliemi H., Tromp G., Prockop D.J.;
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT associated collagen (type IX), and network-forming collagen (type X)
RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL Hum. Mutat. 9:300-315(1997).
RN [12]
RP REVIEW ON OI VARIANTS.
RX MEDLINE=91374476; PubMed=1895312;
RA Byers P.H., Wallis G.A., Willing M.C.;
RT "Osteogenesis imperfecta: translation of mutation to phenotype.";
RL J. Med. Genet. 28:433-442(1991).
RN [13]
RP REVIEW ON OI VARIANTS.
RX MEDLINE=97169389; PubMed=9016532;
RA Daigleish R.;
RT "The human type I collagen mutation database.";
RL Nucleic Acids Res. 25:181-187(1997).
RN [14]
RP VARIANT OI-II CYS-1166.
RX MEDLINE=86287390; PubMed=3016737;
RA Conn D.H., Byers P.H., Steinmann B., Gelinas R.E.;
RT "Lethal osteogenesis imperfecta resulting from a single nucleotide
RT change in one human pro alpha 1(I) collagen allele.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:6045-6047(1986).
RN [15]
RP VARIANT OI-II CYS-569.
RX MEDLINE=87222295; PubMed=3108247;
RA Bateman J.F., Chan D., Walkers I.D., Rogers J.G., Cole W.G.;
RT "Lethal perinatal osteogenesis imperfecta due to the substitution of
RT arginine for glycine at residue 391 of the alpha 1(I) chain of type I
RT collagen.";
RL J. Biol. Chem. 262:7021-7027(1987).
RN [16]
RP VARIANT OI-II CYS-926.

RX MEDLINE=88033031; PubMed=3667599;
RA Vogel B.E., Minor R.R., Freund M., Prockop D.J.;
RT "A point mutation in a type I procollagen gene converts glycine 748
RT of the alpha 1 chain to cysteine and destabilizes the triple helix in
RT a lethal variant of osteogenesis imperfecta.";
RL J. Biol. Chem. 262:14737-14744(1987).
RN [17]
RP VARIANT OI-II ARG-842.
RX MEDLINE=88298828; PubMed=3403550;
RA Bateman J.F., Lamande S.R., Dahl H.H., Chan D., Cole W.G.;
RT "Substitution of arginine for glycine 684 in the collagen alpha 1(I)
RT chain in lethal perinatal osteogenesis imperfecta. Demonstration of
RT the peptide defect by in vitro expression of the mutant cDNA.";
RL J. Biol. Chem. 263:11627-11630(1988).
RN [18]
RP VARIANT OI CYS-1195.
RX MEDLINE=89218628; PubMed=3244312;
RA Labhard M.E., Wirtz M.K., Pope F.M., Nicholls A.C., Hollister D.W.;
RT "A cysteine for glycine substitution at position 1017 in an alpha
RT 1(I) chain of type I collagen in a patient with mild dominantly
RT inherited osteogenesis imperfecta.";
RL Mol. Biol. Med. 5:197-207(1988).
RN [19]
RP VARIANT OI-II VAL-434.
RX MEDLINE=89255493; PubMed=2470760;
RA Patterson E., Smiley E., Boradio J.;
RT "RNA sequence analysis of a perinatal lethal osteogenesis imperfecta
RT mutation.";
RL J. Biol. Chem. 264:10083-10087(1989).
RN [20]
RP VARIANT OI-IV SER-1010.
RX MEDLINE=89308591; PubMed=2745420;
RA Marini J.C., Grange D.K., Gottesman G.S., Lewis M.B., Koeplin D.A.;
RT "Osteogenesis imperfecta type IV. Detection of a point mutation in
RT one alpha 1(I) collagen allele (COL1A1) by RNA/RNA hybrid analysis.";
RL J. Biol. Chem. 264:11893-11900(1989).
RN [21]
RP VARIANTS OI-II ALA-1106, VAL-1151, ARG-1154 AND VAL-1184.
RX MEDLINE=89380165; PubMed=2777764;
RA Lamande S.R., Dahl H.-H.M., Cole W.G., Bateman J.F.;
RT "Characterization of point mutations in the collagen COL1A1 and
RT COL1A2 genes causing lethal perinatal osteogenesis imperfecta.";
RL J. Biol. Chem. 264:15809-15812(1989).
RN [22]
RP VARIANT OI SER-1022.
RX MEDLINE=90062068; PubMed=2511192;
RA Pack M., Constantinou C.D., Kalia K., Nielsen K.B., Prockop D.J.;
RT "Substitution of serine for alpha 1(I)-glycine 844 in a severe
RT variant of osteogenesis imperfecta minimally destabilizes the triple
RT helix of type I procollagen. The effects of glycine substitutions on
RT thermal stability are either position of amino acid specific.";
RL J. Biol. Chem. 264:19694-19699(1989).
RN [23]
RP VARIANT OI-II CYS-1082.
RX MEDLINE=89109573; PubMed=2913053;
RA Constantinou C.D., Nielsen K.B., Prockop D.J.;
RT "A lethal variant of osteogenesis imperfecta has a single base
RT mutation that substitutes cysteine for glycine 904 of the alpha 1(I)
RT chain of type I procollagen. The asymptomatic mother has an
RT unidentified mutation producing an overmodified and unstable type I
RT procollagen.";
RL J. Clin. Invest. 83:574-584(1989).
RN [24]
RP VARIANTS OI CYS-272, CYS-704 AND CYS-896.
RX MEDLINE=90009313; PubMed=2794057;
RA Starman B.J., Eyre D., Charbonneau H., Harrylock M., Weiss M.A.,
RA Weiss L., Graham J.M., Byers P.H.;
RT "Osteogenesis imperfecta. The position of substitution for glycine by
RT cysteine in the triple helical domain of the pro alpha 1(I) chains of
RT type I collagen determines the clinical phenotype.";
RL J. Clin. Invest. 84:1206-1214(1989).
RN [25]
RP VARIANT OI-II CYS-422.

Qy 241 GPGAVGVPAGKDEAGAGGPPGAGPAGGERGEPAGSGFEGLPAGPPGAGKPGGE 300
Db 591 GPGAVGVPAGKDEAGAGGPPGAGPAGGERGEPAGSGFEGLPAGPPGAGKPGGE 650
Qy 301 GVPDGLGAPGSGARGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 360
Db 651 GVPDGLGAPGSGARGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 710
Qy 361 GAGLEGNPGERGAAGLPBPKBDRGDGPKGADGSPKDKVRLTGPIGPPGAGK 420
Db 711 GAGLQMPGERGAAGLPBPKBDRGDGPKGADGSPKDKVRLTGPIGPPGAGK 770
Qy 421 GESPPSPAGTARGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 480
Db 771 GEAGSPSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 830
Qy 481 GPAGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 540
Db 831 GPAGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 890
Qy 541 GKEG 544
Db 891 GKEG 894
RESULT 4
CALL CHICK
ID -CALL CHICK STANDARD; PRT; 1453 AA.
AC P02457;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN COL1A1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE OF 1-153 FROM N.A.
RX MEDLINE=89056316; PubMed=3678834;
RA Finer M.H., Boedtker H., Doty P.;
RT "Construction and characterization of cDNA clones encoding the 5' end
of the chicken pro alpha 1(I) collagen mRNA.";
RL Gene 56:71-78(1987).
RN [2]
RP SEQUENCE OF 1-144 FROM N.A.
RX MEDLINE=88007542; PubMed=2820966;
RA Finer M.H., Aho S., Gerstenfeld L.C., Boedtker H., Doty P.;
RT "Unusual DNA sequences located within the promoter region and the
first intron of the chicken pro-alpha 1(I) collagen gene.";
RL J. Biol. Chem. 262:13323-13332(1987).
RN [3]
RP SEQUENCE OF 152-1187.
RX MEDLINE=82231395; PubMed=7093229;
RA Higberger J.H., Corbett C., Dixit S.N., Yu W., Seyer J.M.,
Kang A.H., Gross J.;
RT "Amino acid sequence of chick skin collagen alpha 1(I)-C8S and the
complete primary structure of the helical portion of the chick skin
collagen alpha 1(I) chain.";
RL Biochemistry 21:2048-2055(1982).
RN [4]
RP SEQUENCE OF 1200-1205.
RX MEDLINE=72243016; PubMed=5047697;
RA Eyre D.R., Glimcher M.J.;
RT "Evidence for a previously undetected sequence at the carboxyterminus
of the alpha 1 chain of chicken bone collagen.";
RL Biochem. Biophys. Res. Commun. 48:720-726(1972).
RN [5]
RP SEQUENCE OF 981-1453 FROM N.A.
RX MEDLINE=81160715; PubMed=6927845;
RA Fuller F., Boedtker H.;

"Sequence determination and analysis of the 3' region of chicken pro-
alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids
including the carboxy-terminal propeptide sequences.";
RL Biochemistry 20:996-1006(1981).
RN [6]
RP SEQUENCE OF 1311-1453 FROM N.A.
RX MEDLINE=80134546; PubMed=6987088;
RA Shewalter A.M., Pesciotta D.M., Eikenberry E.F., Yamamoto T.,
Pastan I., Decrombrughe B., Fietzek P.P., Olsen B.R.;
RT "Nucleotide sequence of a collagen cDNA-fragment coding for the
carboxyl end of pro alpha 1(I)-chains.";
RL FEBS Lett. 111:61-65(1980).
CC -!- FUNCTION: Type I collagen is a member of group I collagen
(fibrillar forming collagen).
CC -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC -!- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
bones. In bones the fibrils are mineralized with calcium
hydroxyapatite.
CC -!- PTM: Proline residues at the third position of the tripeptide
repeating unit (G-X-Y) are hydroxylated in some or all of the
chains. Pro-1153 is the only 3-hydroxypro and the only
hydroxylated proline in position X.
CC -!- SIMILARITY: Contains 1 VWFC domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC
CC EMBL; M17839; AAA48704.1; -
CC EMBL; M17838; AAA48704.1; JOINED.
CC EMBL; V00401; CAA23695.1; -
CC EMBL; M10571; AAA48671.1; ALT_SEQ.
CC EMBL; M17607; AAA48672.1; -
CC PIR; A27179; A27179.
CC PIR; I50629; I50629.
CC InterPro; IPR008161; Clg_helix.
CC InterPro; IPR008160; Collagen.
CC InterPro; IPR000885; Fib_collagen_C.
CC InterPro; IPR002181; Fibrinogen_C.
CC InterPro; IPR001007; VWFC.
CC Pfam; PF01410; COLFI; 1.
CC Pfam; PF01391; Collagen; 18.
CC Pfam; PF00093; vwc; 1.
CC ProDom; PD000007; Clg_helix; 2.
CC ProDom; PD002078; Fib_collagen_C; 1.
CC SMART; SM00038; COLFI; 1.
CC SMART; SM00214; VWFC; 1.
CC PROSITE; PS01208; VWFC_1; 1.
CC PROSITE; PS0184; VWFC_2; 1.
CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Signal; Pyrrolidone carboxylic acid.
FT SIGNAL 1 22
FT PROPEP 23 151 AMINO-TERMINAL PROPEPTIDE.
FT CHAIN 152 1205 COLLAGEN ALPHA 1(I) CHAIN.
FT PROPEP 1206 1453 C-TERMINAL PROPEPTIDE.
FT DOMAIN 31 89 VWFC.
FT MOD_RES 152 152 PYRROLIDONE CARBOXYLIC ACID.
FT MOD_RES 254 254 HYDROXYLATION (POTENTIAL).
FT MOD_RES 851 851 HYDROXYLATION (POTENTIAL).
FT MOD_RES 1081 1081 HYDROXYLATION (POTENTIAL).
FT MOD_RES 1097 1097 HYDROXYLATION (POTENTIAL).
FT MOD_RES 1153 1153 HYDROXYLATION.
FT MOD_RES 1187 1187 F -> L (IN REF. 5).
FT CONFLICT 1187 1187 Q -> H (IN REF. 6).
FT CONFLICT 1441 1441 Q -> H (IN REF. 6).
SQ SEQUENCE 1453 AA; 137789 MW; 3BC6152134271F4D CRC64;
Query Match 90.4%; Score 2775; DB 1; Length 1453;
Best Local Similarity 89.0%; Pred. No. 1.2e-105;
Matches 484; Conservative 28; Mismatches 32; Indels 0; Gaps 0;

SEQUENCE OF 963-1418 FROM N.A.
MEDLINE=95190534; PubMed=3857598;
Cheah K.S.E., Stoker N.G., Griffin J.R., Grosveld F.G., Solomon E.;
"Identification and characterization of the human type II collagen
gene (COL2A1).";
Proc. Natl. Acad. Sci. U.S.A. 82:2555-2559(1985).
[5]
SEQUENCE OF 1120-1398 FROM N.A.
MEDLINE=95306861; PubMed=3840017;
Elima K., Maekelae J.K., Vuorio T., Kauppinen S., Knowles J.,
Vuorio E.;
"Construction and identification of a cDNA clone for human type II
procollagen mRNA.";
Biochem. J. 229:183-188(1985).
[6]
SEQUENCE OF 1106-1418 FROM N.A.
MEDLINE=88067771; PubMed=2825137;
Elima K., Vuorio T., Vuorio E.;
"Determination of the single polyadenylation site of the human pro
alpha 1(II) collagen gene.";
Nucleic Acids Res. 15:9499-9504(1987).
[7]
SEQUENCE OF 1227-1289 FROM N.A.
MEDLINE=86104139; PubMed=3002437;
Nunez A.M., Francomano C., Young M.F., Martin G.R., Yamada Y.;
"Isolation and partial characterization of genomic clones coding for
a human pro-alpha 1 (II) collagen chain and demonstration of
restriction fragment length polymorphism at the 3' end of the gene.";
Biochemistry 24:6343-6348(1985).
[8]
SEQUENCE OF 1176-1226 FROM N.A.
MEDLINE=84118798; PubMed=6320112;
Strom C.M., Upholt W.B.;
"Isolation and characterization of genomic clones corresponding to
the human type II procollagen gene.";
Nucleic Acids Res. 12:1025-1038(1984).
[9]
SEQUENCE OF 35-167 FROM N.A.
MEDLINE=89233138; PubMed=2714801;
Su M.W., Benson-Charda V., Vissing H., Ramirez F.;
"Organization of the exons coding for pro alpha 1(II) collagen N-
propeptide confirms a distinct evolutionary history of this domain of
the fibrillar collagen genes.";
Genomics 4:438-441(1989).
[10]
REVIEW ON VARIANTS.
MEDLINE=91184577; PubMed=2010058;
Kuivaniemi H., Tromp G., Prockop D.J.;
"Mutations in collagen genes: causes of rare and some common diseases
in humans.";
FASEB J. 5:2052-2060(1991).
[11]
REVIEW ON VARIANTS.
MEDLINE=97255959; PubMed=9101290;
Kuivaniemi H., Tromp G., Prockop D.J.;
"Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
associated collagen (type IX), and network-forming collagen (type X)
cause a spectrum of diseases of bone, cartilage, and blood vessels.";
Hum. Mutat. 9:300-315(1997).
[12]
VARIANT SER-1074.
MEDLINE=90036909; PubMed=2572591;
Vissing H., D'Alessio M., Lee B., Ramirez F., Godfrey M.,
Hollister D.W.;
"Glycine to serine substitution in the triple helical domain of pro-
alpha 1 (II) collagen results in a lethal perinatal form of short-
limbed dwarfism";
J. Biol. Chem. 264:18265-18267(1989).
[13]
VARIANT SEDC 1095-GLY--TYR-1330 DEL.
MEDLINE=89266907; PubMed=2543071;
Lee B., Vissing H., Ramirez F., Rogers D., Rimoin D.;
"Identification of the molecular defect in a family with
short-limbed dwarfism";

propeptide confirms a distinct evolutionary history of this domain of the fibrillar collagen genes.;"

[10] Genomics 4:438-441(1989).

NP REVIEW ON VARIANTS.

XX MEDLINE=91184577; PubMed=2010058;

XX Kuivaniemi H., Tromp G., Prockop D.J.;

XX "Mutations in collagen genes: causes of rare and some common diseases in humans.;"

XX FASEB J. 5:2052-2060(1991).

XX [11]

NP REVIEW ON VARIANTS.

XX MEDLINE=97253959; PubMed=9101290;

XX Kuivaniemi H., Tromp G., Prockop D.J.;

XX "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-associated collagen (type IX), and network-forming collagen (type X) cause a spectrum of diseases of bone, cartilage, and blood vessels.;"

XX Hum. Mutat. 9:300-315(1997).

XX [12]

NP VARIANT SER-1074.

XX MEDLINE=90036909; PubMed=2572591;

XX Vissing H., D'Alessio M., Lee B., Ramirez F., Godfrey M., Hollister D.W.;

XX "Glycine to serine substitution in the triple helical domain of pro-alpha 1 (I) collagen results in a lethal perinatal form of short-limbed dwarfism.;"

XX J. Biol. Chem. 264:18265-18267(1989).

XX [13]

NP VARIANT SEDC 1095-GLY--TYR-1330 DEL.

XX MEDLINE=89266907; PubMed=2543071;

XX Lee B., Vissing H., Ramirez F., Rogers D., Rimoin D.;

XX "Identification of the molecular defect in a family with

RT spondyloepiphyseal dysplasia.";
 RL Science 244:978-980(1989).
 RA [14]
 RP VARIANT OSTEOARTHRTIS CYS-650.
 RX MEDLINE=90370826; PubMed=1975693;
 RA Ala-Kokko L., Baldwin C.T., Moskowitz R.W., Prockop D.J.;
 RT "Single base mutation in the type II procollagen gene (COL2A1) as a
 RT cause of primary osteoarthritis associated with a mild
 RT chondrodysplasia.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:6565-6568(1990).
 RN [15]
 RP VARIANT OI-IV VAL-717.
 RX MEDLINE=91291136; PubMed=2064612;
 RA Bateman J.F., Hannagan M., Chan D., Cole W.G.;
 RT "Characterization of a type I collagen alpha 2(I) glycine-586 to
 RT valine substitution in osteogenesis imperfecta type IV. Detection of
 RT the mutation and prenatal diagnosis by a chemical cleavage method.";
 RL Biochem. J. 276:765-770(1991).
 RN [16]
 RP VARIANT OSTEOARTHRTIS CYS-650.
 RX MEDLINE=91086471; PubMed=1985108;
 RA Eyre D.R., Weis M.A., Moskowitz R.W.;
 RT "Cartilage expression of a type II collagen mutation in an inherited
 RT form of osteoarthritis associated with a mild chondrodysplasia.";
 RL J. Clin. Invest. 87:357-361(1991).
 RN [17]
 RP VARIANT HYPOCHONDROGENESIS GLU-984.
 RX MEDLINE=93054548; PubMed=1429602;
 RA Bogaert R., Tiller G.E., Wies M.A., Gruber H.E., Rimoin D.L.,
 RA Cohn D.H., Eyre D.R.;
 RT "An amino acid substitution (Gly853-->Glu) in the collagen alpha
 RT 1(II) chain produces hypochondrogenesis.";
 RL J. Biol. Chem. 267:22522-22526(1992).
 RN [18]
 RP VARIANT HYPOCHONDROGENESIS SER-705.
 RX MEDLINE=92262484; PubMed=1374906;
 RA Horton W.A., Machado M.A., Ellard J., Campbell D., Bartley J.,
 RA Ramirez F., Vitale E., Lee B.;
 RT "Characterization of a type II collagen gene (COL2A1) mutation
 RT identified in cultured chondrocytes from human hypochondrogenesis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:4583-4587(1992).
 RN [19]
 RP VARIANT WS-II ASP-198.
 RX MEDLINE=93304428; PubMed=8317498;
 RA Koerkoe J., Ritvaniemi P., Haataja L., Kaaerlaeinen H.,
 RA Kivirikko K.I., Prockop D.J., Ala-Kokko L.;
 RT "Mutation in type II procollagen (COL2A1) that substitutes aspartate
 RT for glycine alpha 1-67 and that causes cataracts and retinal
 RT detachment: evidence for molecular heterogeneity in the Wagner
 RT syndrome and the Stickler syndrome (arthro-ophthalmopathy).";
 RL Am. J. Hum. Genet. 53:55-61(1993).
 RN [20]
 RP VARIANT SEMD CYS-840.
 RA Tiller G.E., Weis M.A., Lachman R.S., Cohn D.H., Rimoin D.L.,
 RA Eyre D.R.;
 RT "A dominant mutation in the type II collagen gene (COL2A1) produces
 RT spondyloepimetaphyseal dysplasia (SEMD), Strudwick type.";
 RL Am. J. Hum. Genet. 53:A209-A209(1993).
 RN [21]
 RP VARIANT OSTEOARTHRTIS CYS-650.
 RX MEDLINE=93282819; PubMed=8507190;
 RA Holderbaum D., Maledum C.J., Moskowitz R.W., Haqqi T.M.;
 RT "Human cartilage from late stage familial osteoarthritis transcribes
 RT type II collagen mRNA encoding a cysteine in position 519.";
 RL Biochem. Biophys. Res. Commun. 192:1169-1174(1993).
 RN [22]
 RP VARIANT SEMD ARG-285.
 RX MEDLINE=93252400; PubMed=8486375;
 RA Vakkula M., Ritvaniemi P., Vuorio A.P., Ala-Kokko L.,
 RA Peltonen L.;
 RT "A mutation in the amino-terminal end of the triple helix of type II
 RT collagen causing severe osteochondrodysplasia.";
 RL Genomics 16:282-285(1993).

RN [23]
 RP VARIANT SEMD CYS-206.
 RX MEDLINE=94063862; PubMed=8244341;
 RA Williams C.J., Considine E.L., Knowlton R.G., Reginato A., Neumann G.,
 RA Harrison D., Buxton P., Jimenez S.A., Prockop D.J.;
 RT "Spondyloepiphyseal dysplasia and precocious osteoarthritis in a
 RT family with an Arg75-->Cys mutation in the procollagen type II gene
 RT (COL2A1).";
 RL Hum. Genet. 92:499-505(1993).
 RN [24]
 RP VARIANT SEMD CYS-920.
 RX MEDLINE=93315508; PubMed=8325895;
 RA Chan D., Taylor T.K.F., Cole W.G.;
 RT "Characterization of an arginine 789 to cysteine substitution in
 RT alpha 1 (II) collagen chains of a patient with spondyloepiphyseal
 RT dysplasia.";
 RL J. Biol. Chem. 268:15238-15245(1993).
 RN [25]
 RP VARIANT SEMD SER-1128.
 RX MEDLINE=93140139; PubMed=8423604;
 RA Cole W.G., Hall R.K., Rogers J.G.;
 RT "The clinical features of spondyloepiphyseal dysplasia congenita
 RT resulting from the substitution of glycine 997 by serine in the alpha
 RT 1(II) chain of type II collagen.";
 RL J. Med. Genet. 30:27-35(1993).
 / Query Match 71.7%; Score 2202; DB 1; Length 1418;
 / Best Local Similarity 71.9%; Pred. No. 1.3e-82;
 / Matches 391; Conservative 40; Mismatches 113; Indels 0; Gaps 0;
 QY 1 GSEGGVGVGPGPGPAGAGPAGDPGADGPGAGKADGAPGIAGAPGPGARGSPGE 60
 Db 315 GPEGAGPGEPTGTPSPAGASGNPTGDIPIGAKGSAGAGIAGAPGPGRGPDPQ 374
 QY 61 GPGGPFPGKDSGEPGAPGSKGDTAKGPPGVGVGPPGPPGAGEGKPGARPGTGLP 120
 Db 375 GATGTLGPKGTGKPGIAGFKGEQPGKGPAGPAGPAGPAGPAGPAGPAGPAGP 434
 QY 121 GPPGEGGSGSGPFGADGAGKPGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 180
 Db 435 GPPGEGGSGSGPFGADGAGKPGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 494
 QY 181 GSPGSPGPDGKTGPPGPGAGDGRPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 240
 Db 495 GRPDAGPQGVGSPGAPGDPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 554
 QY 241 GPPGAGPAGKDGAGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 300
 Db 555 GAFGLRGLPKDGETGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 614
 QY 301 GVPDILCAPGSPGARGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 360
 Db 615 GVPGEAGAPGLVGRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 674
 QY 361 GAFGLEMPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 420
 Db 675 GPPGLQMPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 734
 QY 421 GESGPPSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 480
 Db 735 GEVGPFGPAGSAGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 794
 QY 481 GPAGPAGPFGPIGDVGPAGKAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 540
 Db 795 GPQGPSGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 854
 QY 541 GKGG 544
 Db 855 KGG 858

RESULT 6
 CA12_MOUSE

ID CA12 MOUSE STANDARD; PRT; 1459 AA.
 AC P28481, PRT; 1459 AA.
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Collagen alpha 1(II) chain precursor [Contains: Chondrocalcin].
 GN COL2A1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RX MEDLINE=91358489; PubMed=1885613;
 RA Metaranta M., Toman D., de Crombrughe B., Vuorio E.;
 RT "Mouse type II collagen gene. Complete nucleotide sequence, exon
 structure and alternative splicing."
 RL J. Biol. Chem. 266:16862-16869 (1991).
 RN [2]
 RP SEQUENCE OF 1455-1459 FROM N.A.
 RX MEDLINE=91274355; PubMed=2054384;
 RA Metaranta M., Toman D., de Crombrughe B., Vuorio E.;
 RT "Specific hybridization probes for mouse type I, II, III and IX
 collagen mRNAs."
 RL Biochim. Biophys. Acta 1089:241-243 (1991).
 CC -!- FUNCTION: Collagen type II is specific for cartilaginous tissues.
 CC -!- SUBUNIT: Homotrimers of alpha 1(II) chains.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=Long;
 CC IsoId=28481-1; Sequence=Displayed;
 CC Name=Short;
 CC IsoId=28481-2; Sequence=VSP_001139, VSP_001140;
 CC -!- PRT: Prolines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -!- SIMILARITY: Contains 1 WFC domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M65161; AAA68100.1; -;
 DR EMBL; X57982; CAA41047.1; -;
 DR MGD; MG1:88452; Col2a1.
 DR InterPro; IPR008161; Clq helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR000885; Fib collagen_C.
 DR InterPro; IPR001007; WFC_C.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 18.
 DR Pfam; PF00093; wvc; 1.
 DR ProDom; PD000007; Clq helix; 5.
 DR ProDom; PD002078; Fib collagen_C; 1.
 DR SMART; SM00038; COLFI_1.
 DR SMART; SM00214; WFC; 1.
 DR PROSITE; PS01208; WFC; 1; 1.
 DR PROSITE; PS50184; WFC; 2; 1.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen; Signal; Alternative splicing.
 FT SIGNAL 1 25 POTENTIAL.
 FT PROPEP 26 153 AMINO-TERMINAL PROPEPTIDE
 FT (BY SIMILARITY).
 FT CHAIN 154 1213 COLLAGEN ALPHA 1(II) CHAIN.
 FT PROPEP 1214 1459 CHONDROCALCIN.
 FT DOMAIN 32 89 WFC.
 FT DOMAIN 173 1186 TRIPLE-HELICAL REGION.
 FT DOMAIN 1187 1213 NONHELICAL REGION (C-TERMINAL).
 FT VARSPPLIC 29 29 Q -> R (in isoform Short).
 FT /FTId=VSP_001139.

FT VARSPPLIC 30 98 Missing (in isoform Short).
 FT /FTId=VSP_001140.
 SQ SEQUENCE 1459 AA; 139154 MW; F6C84FA7C532E7F2 CRC64;
 Query Match 71.4%; Score 2192; DB 1; Length 1459;
 Best Local Similarity 71.7%; Pred. No. 3.4e-82;
 Matches 390; Conservative 40; Mismatches 114; Indels 0; Gaps 0;
 QY 1 GSEPGVGRGEPGPPGAGAGAGADPGADGPGAGKAGDAGAGIAGAGFPGAGPSPGPE 60
 DB 356 GPEGAGSGRGEPCNPGSPAGAGSNPGTDGIPGAKGSAGAGIAGAGFPGPRGPPGQ 415
 QY 61 GPGGPPGKGDSEPCAPGSKGDTGAKGEPGVGVGPPGPPGAGEGKGGKAGEGPGTGLP 120
 DB 416 GATGPIGPKGQAGEPGIAGFGKDGQPKGETGAPGQAGPAGBEGKRGKAGEPGGAGPI 475
 QY 121 GPPGPGSGRSGFTGADGVAGPKGAPGERSGPPGAGPKGSPGEGAGRPCEAGLPKAGLIT 180
 DB 476 GPPGGERGAPGNRGFTGQDGLAGPKGAPGERGSPGLAGFKGANGDPRGEPCLPGARGLT 535
 QY 181 GSPGSGPPDGKTPPGPAGEDDRPPGPPGARGAGVMGPPGKGAAGEFGKAGRGVP 240
 DB 536 GRPGDAGPGQKVGPSGAFGEDEGRPPGPGQARGQPGVMGPPGKANGGEFGKAGEKGLA 595
 QY 241 GPPGAVGPAKDGEGAGAGPPGPPGAPGAGERGEGPAGSPGFEGLPGPAGPGEAGKPGEE 300
 DB 596 GAGPLRGLPKDGETGAAGPPGPPGSPGAGERGEGQAGPSPGQGLPGPPGPGEGGKQGDQ 655
 QY 301 GVPDGLAPGSGARGEPGPPGGERVVEGPPGADGAPGDDGAKGDAGAPGAGE 360
 DB 656 GIPGEGAPGLVGRGERGFTGERSGPAQGLQGPRGLFTGTPGTGPKAAGAPGPPGCAQ 715
 QY 361 GAGPLEGMPGERRAAGLPKGDGADGKADGSPGKDGVRGLTGTGTPGPPGAGAGDK 420
 DB 716 GPFGLQGVGREGAAGIAGPKGDRGDKGKPEAGPKDGGRLTGTGTPGPPGAGANGEK 775
 QY 421 GBSGSGPAGTGTGARGAFDGRGEPGPPGPPGAGPPGADGEPGAKGPPGDAGKDGAGPP 480
 DB 776 GEGAGPPSGSGTGARGAGPEGETGPPGPPGAGPPGADGQPGAKGDOEGAGKGDAGAP 835
 QY 481 GPAGAGPPGPIGVGAPKAGSGSAGPCATGPPGAGRGVGGPPGSGDAGPPGPPGPA 540
 DB 836 GPGGSGAPPGQGTGVTGPKARGAQGPATGTPGGAAGRVGPPGANGNPPGPPGPA 895
 QY 541 GKEX 544
 DB 896 GKDX 899
 RESULT 7
 CALL RAT STANDARD; PRT; 671 AA.
 AC P02454; P02455;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Collagen alpha 1(II) chain (fragments).
 GN COL1A1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE OF 1-19.
 RX MEDLINE=69155173; PubMed=5777344;
 RA Bornstein P.;
 RT "Comparative sequence studies of rat skin and tendon collagen. II.
 RT The absence of a short sequence at the amino terminus of the skin
 RT alpha-1 chain."
 RL Biochemistry 8:63-71 (1969).
 RN [2]
 RP SEQUENCE OF 5-19.
 RX MEDLINE=67162268; PubMed=5337886;

RESULT 9

ID	CA21_BOVIN	STANDARD;	PRT;	1364 AA.
AC	P02455; 062549;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Collagen alpha 2(I) chain precursor.			
GN	COL1A2.			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Bovinae; Bos.			
OX	NCBI_TaxID=9913;			
OX	[1]			
RN	SEQUENCE FROM N.A.			
RP	TISSUE=Aorta;			
RC	MEDLINE=98290219; PubMed=9628255;			
RA	Shirai T., Hattori S., Sakaguchi M., Inouye S., Kimura A., Ebihara T.,			
RA	Irie S., Nagai Y., Hori H.,			
RA	"The complete cDNA coding sequence for the bovine proalpha2(I) chain			
RT	of type I procollagen.";			
RT	Matrix Biol. 17:85-88(1998).			
RN	[2]			
RN	SEQUENCE OF 80-98.			
RP	TISSUE=Skin;			
RC	MEDLINE=75036115; PubMed=4609475;			
RA	Fietzek P.P., Breitkreutz D., Kuehn K.;			
RA	"Amino acid sequence of the amino-terminal region of calf skin			
RT	collagen.";			
RT	Biochim. Biophys. Acta 365:305-310(1974).			
RN	[3]			
RN	SEQUENCE OF 95-415, AND REVISION.			
RP	TISSUE=Skin;			
RC	MEDLINE=76091874; PubMed=173531;			
RA	Fietzek P.P., Rexrodt F.W.;			
RA	"The covalent structure of collagen. The amino-acid sequence of			
RT	alpha2-CB4 from calf-skin collagen.";			
RT	Eur. J. Biochem. 59:113-118(1975).			
RN	[4]			
RN	SEQUENCE OF 416-445.			
RP	TISSUE=Skin;			
RC	MEDLINE=75008198; PubMed=4412529;			
RA	Fietzek P.P., Furchmayr H., Kuehn K.;			
RA	"Comparative sequence studies on alpha2-CB2 from calf, human, rabbit			
RT	and pig-skin collagen.";			
RT	Eur. J. Biochem. 47:257-261(1974).			
RN	[5]			
RN	SEQUENCE OF 446-481.			
RP	TISSUE=Skin;			
RC	MEDLINE=75059250; PubMed=4435743;			
RA	Fietzek P.P., Kuehn K.;			
RA	"The covalent structure of collagen: amino acid sequence of the N-			
RT	terminal region of alpha2-CB3 from rat skin collagen and alpha2-CB3.5			
RT	from calf skin collagen.";			
RT	Hoppe-Seyler's Z. Physiol. Chem. 355:647-650(1974).			
CC	-1- FUNCTION: Type I collagen is a member of group I collagen			
CC	(fibrillar forming collagen).			
CC	-1- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.			
CC	-1- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and			
CC	bones. In bones the fibrils are mineralized with calcium			
CC	hydroxapatite.			
CC	-1- PTM: Prolines at the third position of the tripeptide repeating			
CC	unit (G-X-Y) are hydroxylated in some or all of the chains.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			


```

QY 532 GPPGPPGAGKEG 544
DB 899 GPPGHPGAGNG 911

RESULT 11
CA21_RAT
ID CA21_RAT STANDARD; PRT; 1372 AA.
AC P02466; Q9R188;
DT 21-JUL-1986 (Rel. 01, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 2(I) chain precursor.
GN COL1A2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Guenther D., Seibold S., Marx M.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 86-98.
RC TISSUE=Skin;
RX MEDLINE=67162268; PubMed=5337886;
RA Kang A.H., Bornstein P., Piez K.A.;
RT "The amino acid sequence of peptides from the cross-linking region of
RT rat skin collagen."
RL Biochemistry 6:788-795 (1967).
RN [3]
RP SEQUENCE OF 99-102.
RC TISSUE=Skin;
RX MEDLINE=69206881; PubMed=5785232;
RA Fietzek P.P., Piez K.A.;
RT "Isolation and characterization of the cyanogen bromide peptides from
RT the alpha 2 chain of rat skin collagen."
RL Biochemistry 8:2129-2133 (1969).
RN [4]
RP SEQUENCE OF 102-144.
RC TISSUE=Skin;
RX MEDLINE=73049496; PubMed=4636752;
RA Fietzek P.P., Kell I., Kuehn K.;
RT "The covalent structure of collagen. Amino acid sequence of the N-
RT terminal region of alpha 2-CB4 from calf and rat skin collagen."
RL FEBS Lett. 26:66-68 (1972).
RN [5]
RP SEQUENCE OF 423-452.
RC TISSUE=Skin;
RX MEDLINE=71115216; PubMed=5544553;
RA Hightberger J.H., Kang A.H., Gross J.;
RT "Comparative studies on the amino acid sequence of the alpha 2-CB2
RT peptides from chick and rat skin collagens."
RL Biochemistry 10:610-616 (1971).
RN [6]
RP SEQUENCE OF 453-501.
RC TISSUE=Skin;
RX MEDLINE=75059250; PubMed=4435743;
RA Fietzek P.P., Kuehn K.;
RT "The covalent structure of collagen: amino acid sequence of the N-
RT terminal region of alpha2-CB3 from rat skin collagen and alpha2-CB3.5
RT from calf skin collagen."
RL Hoppe-Seyler's Z. Physiol. Chem. 355:647-650 (1974).
RN [7]
RP SEQUENCE OF 791-836.
RC TISSUE=Skin;
RX MEDLINE=74055004; PubMed=4763308;
RA Fietzek P.P., Kuehn K.;
RT "The covalent structure of collagen: amino acid sequence of the N-
RT terminal region of alpha 2-CB5 from rat skin collagen."
RL FEBS Lett. 36:289-291 (1973).
RN [8]

```

```

RP ORDER OF CNBR PEPTIDES.
RX MEDLINE=70181852; PubMed=5443712;
RA Vuust J., Lane J.M., Fietzek P.P., Miller E.J., Piez K.A.;
RT "The order of the CNBR peptides from the alpha 2 chain of collagen."
RL Biochem. Biophys. Res. Commun. 38:703-708 (1970).
CC -I- FUNCTION: Type I collagen is a member of group I collagen
CC (fibrillar forming collagen).
CC -I- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC -I- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
CC bones. In bones the fibrils are mineralized with calcium
CC hydroxyapatite.
CC -I- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF121217; AAD41775.1; -.
DR InterPro: IPR008161; Clg_helix.
DR InterPro: IPR008160; Collagen.
DR InterPro: IPR000885; Fib_collagen_C.
DR Pfam: PF01410; COLFI; 1.
DR Pfam: PF01391; Collagen; 18.
DR ProDom: PD000007; Clg_helix; 4.
DR ProDom: PD002078; Fib_collagen_C; 1.
DR SMART: SM00038; COLFI; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Signal; Pyrrolidone carboxylic acid.
FT SIGNAL 1 24 POTENTIAL.
FT PROPEP 25 85 AMINO-TERMINAL PROPEPTIDE
FT CHAIN 86 1108 (BY SIMILARITY).
FT PROPEP 1109 1372 COLLAGEN ALPHA 2(I) CHAIN
FT SITE 783 785 CARBOXYL-TERMINAL PROPEPTIDE
FT SITE 828 830 (BY SIMILARITY).
FT SITE 1011 1013 CELL ATTACHMENT SITE (POTENTIAL).
FT MOD_RES 86 86 CELL ATTACHMENT SITE (POTENTIAL).
FT MOD_RES 90 90 PYRROLIDONE CARBOXYLIC ACID (PROBABLE).
FT CARBOHYD 1273 1273 INVOLVED IN CROSS-LINKING.
FT CONFLICT 132 132 N-LINKED (GLCNAC...) (POTENTIAL).
FT CONFLICT 137 137 T -> P (IN REF. 4).
FT CONFLICT 145 422 S -> P (IN REF. 4).
FT CONFLICT 431 432 MISSING (IN REF. 4).
FT CONFLICT 431 432 ST -> TS (IN REF. 5).
FT CONFLICT 494 494 E -> Z (IN REF. 6).
FT CONFLICT 497 497 N -> A (IN REF. 6).
FT CONFLICT 502 790 MISSING (IN REF. 6).
FT CONFLICT 825 825 R -> K (IN REF. 7).
SQ SEQUENCE 1372 AA; 129564 MW; B069371A8DB20A72 CRC64;

Query Match 63.0%; Score 1934; DB 1; Length 1372;
Best Local Similarity 64.7%; Pred. No. 7,8e-72;
Matches 352; Conservative 44; Mismatches 148; Indels 0; Gaps 0;

QY 1 GSEGGVGRGEPGPPGAGAGPAGDPGADGPGAGKAGDAGFAGIAGAGPFGAGSGPE 60
DB 280 GPAGPAGPGEAGLPLGLSGPVGPPGNPGANGLTGAKGATGLPGVAGAPGLPGRGIPGV 339
QY 61 GPGGPPGKDGSGEPGAPGSKGDTGAKGEPGVPVGVEGPPGAGGKPGARGEPGTGLP 120
DB 340 GAAGATGPRGLVGBEPGAGSKGTGNKGEPGSAGAQGPPGSGEGKRGSPGSGAGPA 399
QY 121 GPPGEGGSGSGFGPGADGAVAGFKGPPAGSRGSPGAPGKSGPGEAGRPGEAGLPKAGLT 180
DB 400 GPPGLRGSFGSRGLPLGADGAGVMGPPGNRGSTGAGVGGPNGDAGRPGEPLMGPRGLP 459
QY 181 GSPGSPGPGDKTGPPGPGAGEDGRGPPGPPGARGAGVGMFFGPKGAAGEGPKAGERGV 240

```

Db 460 GSPGNVGPAGKEGVGLFDGDRPGTGPAGPGEAGNIGFPGPKGSGDPGKPGSKGHP 519
Qy 241 GPPGAVGPAGKDGAGAGEGPPGAPGAGGEEGPGAGSPGFEGLPGPAGPAGGAGKPGEE 300
Db 520 GLACGARGPFGDGNAGQGPFGVQVQKGEQGPAGFPFGQLPGPSGTAGEVGEGER 579
Qy 301 GVPGLGAPGSPGARGPFGPFGGPGVGGPPGAGPAGDAGDAGDAGKAGDAGAPGSE 360
Db 580 GLPGEFLGPGAPGPRGERPFGGPGGSAAGSPGPIGIRGSPGAPGPGNKGAGAVGAPGA 639
Qy 361 GAPLEGMPGPRGAAGLPKPGKDRDAGPKGADGSPKDGVRGLTGPIGPPGAGAPGDK 420
Db 640 CASFGGLPGRGAAGLPKPGKGEKGTGLRGEIGNFGRDARGAPGAGPAGASGR 699
Qy 421 GESFGSPAGTGTARGAPGDRGEPGPPGAGFAGPAGDGEPAKGEPPGAGAKGDPAGPP 480
Db 700 GEAGAAGSPGAPGPRGSPGERGEPGAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 759
Qy 481 GPAGPAGPPGPIGVGAPGAKGARGSPGATGTFPCAAGRVGPPGSGDAGPPGPPGA 540
Db 760 GPTGVGAAGSPGNGPPGPPGAGSKRGDPGPGMTGTPGAGRTGPPGSGITGPPGPPGA 819
Qy 541 GKEG 544
Db 820 GKEG 823

RESULT 12

CA25 HUMAN STANDARD; PRT; 1496 AA.
AC P05997;
DC 01-APR-1988 (Rel. 07, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 2(V) chain precursor.
GN COL5A2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE OF 1-463 FROM N.A.
RA MEDLINE=89123368; PubMed=2914927;
RX Woodbury D., Benson-Chanda V., Ramirez F.;
RT "Amino-terminal propeptide of human pro-alpha 2(V) collagen conforms
to the structural criteria of a fibrillar procollagen molecule.";
RL J. Biol. Chem. 264:2735-2738(1989).
RN [2]
RP SEQUENCE OF 398-1496 FROM N.A.
RA MEDLINE=87146331; PubMed=3029669;
RX Weil D., Bernard M.P., Gargano S., Ramirez F.;
RT "The pro alpha 2(V) collagen gene is evolutionarily related to the
major fibrillar-forming collagens.";
RL Nucleic Acids Res. 15:181-198(1987).
RN [3]
RP SEQUENCE OF 1227-1496 FROM N.A.
RA MEDLINE=85289337; PubMed=2411731;
RX Myers J.C., Loidl H.R., Seyer J.M., Dion A.S.;
RT "Complete primary structure of the human alpha 2 type V procollagen
COOH-terminal propeptide.";
RL J. Biol. Chem. 260:11216-11222(1985).
RN [4]
RP SEQUENCE OF 1449-1496 FROM N.A.
RA MEDLINE=89138450; PubMed=3224983;
RX Tsipouras P., Schwartz R.C., Liddell A.C., Salkeld C.S., Weil D.,
RA Ramirez F.;
RT "Genetic distance of two fibrillar collagen loci, COL3A1 and COL5A2,
located on the long arm of human chromosome 2.";
RL Genomics 3:275-277(1988).
RN [5]
RP SEQUENCE OF 208-227.
RC TISSUE=Placenta;

RX MEDLINE=922339022; PubMed=1571108;
RA Mann K.;
RT "Isolation of the alpha 3-chain of human type V collagen and
characterization by partial sequencing.";
RL Biol. Chem. Hoppe-Seyler 373:69-75(1992).
RN [6]
RP SEQUENCE OF 288-297 AND 606-617.
RC TISSUE=Bone;
RX MEDLINE=94237164; PubMed=8181482;
RA Moradi-Ameli M., Rousseau J.C., Klemm J.P., Champliand M.F.,
RA Boutillon M.M., Bernillon J., Wallach J.M., van der Rest M.;
RT "Diversity in the processing events at the N-terminus of type-V
collagen.";
RL Eur. J. Biochem. 221:987-995(1994).
RN [7]
RP DISEASE.
RX MEDLINE=98087576; PubMed=9425231;
RA Michalikova K., Susic M., Willing M.C., Wenstrup R.J., Cole W.G.;
RT "Mutations of the alpha2(V) chain of type V collagen impair matrix
assembly and produce Ehlers-Danlos syndrome type I.";
RL Hum. Mol. Genet. 7:249-255(1998).
RN [8]
RP VARIANT EDS-II ARG-960.
RX MEDLINE=98455031; PubMed=9783710;
RA Richards A.J., Martin S., Nicholls A.C., Harrison J.B., Pope F.M.,
RA Burrows N.P.;
RT "A single base mutation in COL5A2 causes Ehlers-Danlos syndrome type
II.";
RL J. Med. Genet. 35:846-848(1998).
CC -!- FUNCTION: Type V collagen is a member of group I collagen
(fibrillar forming collagen). It is a minor connective tissue
component of nearly ubiquitous distribution. Type V collagen binds
to DNA, heparan sulfate, thrombospondin, heparin, and insulin.
CC -!- SUBUNIT: Trimers of two alpha 1(V) and one alpha 2(V) chains in
most tissues and trimers of one alpha 1(V), one alpha 2(V), and
one alpha 3(V) chains in placenta.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- DISEASE: Defects in COL5A2 are a cause of Ehlers-Danlos syndrome
type I (EDS-I) [MIM:130000]; also known as Ehlers-Danlos syndrome
gravis. EDS-I is an autosomal dominant connective-tissue disorder
characterized by loose-jointedness and fragile, velvety,
stretchable, bruisable skin that heals with peculiar 'cigarette-
paper' scars. Inheritance is autosomal dominant.
CC -!- DISEASE: Defects in COL5A2 are a cause of Ehlers-Danlos syndrome
type II (EDS-II) [MIM:130010]; also known as Ehlers-Danlos
syndrome mitis. Inheritance is autosomal dominant.
CC -!- SIMILARITY: Contains 1 WFCC domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
or send an email to license@isb-sib.ch).
CC EMBL; J04478; AAA51859.1; -
DR EMBL; X04758; CAA28454.1; -
DR EMBL; M11718; AAA52058.1; -
DR PIR; A31427; CGHU2V.
DR PDB; 1A9A; 18-NOV-98.
DR Genew; HGNC:2210; COL5A2.
DR MIM; 120190; -
DR MIM; 130000; -
DR MIM; 130010; -
DR GO; GO:0005588; C:collagen type V; TAS.
DR GO; GO:0008151; P:cell growth and/or maintenance; TAS.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR002181; Fibrinogen_C.
DR InterPro; IPR001007; VWF_C.

```

DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF00093; vwc; 1.
DR ProDom; PD000007; Cig_helix; 5.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; vwc; 1.
DR PROSITE; PS01208; VWC_1; 1.
DR PROSITE; PS01184; VWC_2; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Signal; Ehlers-Danlos syndrome;
KW Disease mutation; 3D-structure.
FT SIGNAL 1 26
FT CHAIN 27 1226
FT PROPEP 1227 1496
FT DOVAIN 39 97
FT MOD_RES 290 290
FT MOD_RES 293 293
FT MOD_RES 296 296
FT MOD_RES 608 608
FT MOD_RES 614 614
FT VARIANT 960 960
FT CONFLICT 292 292
FT CONFLICT 1418 1418
FT CONFLICT 1438 1438
FT CONFLICT 1460 1460
FT CONFLICT 1496 1496
SQ SEQUENCE 1496 AA; 144720 MW; 82827C7A8644F5A CRC64;

Query Match 62.6%; Score 1921.5; DB 1; Length 1496;
Best Local Similarity 64.0%; Pred. No. 2.6e-71;
Matches 348; Conservative 40; Mismatches 153; Indels 3; Gaps 1;

QY 1 CSEGGVGRGPPGPPGAGAGPAGDPCADCEPGKAGKADGAPGAGPPGARGPSGPE 60
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 396 GPESGQGRGTGPPGVSGPLGALGTDTPPKPTGSGTSGP---PGSAGPPGSP 452
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 GPGGPPGKDSGPPGAPSGKDTGAKGPPGVGVGPPGPPGAGBEGKPGARFPGTGLP 120
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 453 GPQSGTGPGQNSGLPGDPGFGGAGKPGKGPFGHIGQPGIPGPEGKGRGPRGDTGLPP 512
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 121 GPPGEGGPGSGPPGADGVAGPPGAPGERSGPPGAPKSGPSGAGRGEAGLPGAKGLT 180
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 513 GPVGERGAPNGRFGSDGLPFGKAGQGERGPPVSSGPKGSQGDPPGPPGLPGARGLT 572
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 181 GSPGSPGDKGTGPPGAGEDCRPGPPGPPGARGGAGVWGPFGPKGAAGPPGKAGRGVP 240
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 573 GNPVGQGPGLGALGAPGEDRGPPGSGIGIKQPGTMTGLPGPKSGNGDPKPGGAGNP 632
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 241 GPPGAVGPPAGKDGAGAGPPGPPGAPGAGERGEPGAGSPGPEGLPGPAGPPGAGKPGEE 300
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 633 GVPQORGAPGKDGKVGPPGPPGPPGLRGERGEQGPFGPTGFGHPGPPPPGEGGKPGDQ 692
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 301 GVPGLGAPGSGARGEPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 360
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 693 GVPFGGAVGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 752
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 361 GAPLEGMPGPPGAGLPGPKDGRDAGPKGADSPGKDGVRGLTGTPIGPPGAPAGPDK 420
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 753 GPPGLQMPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 812
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 421 GESGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 480
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 813 GEPGRLVPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 872
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 481 GPAGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 540
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 873 GPQGLAGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 932
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 541 GKEG 544
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 933 GKEG 936
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 13

CA21 MOUSE

```

ID CA21 MOUSE STANDARD; PRT; 1372 AA.
AC Q01149;
DT 01-APR-1993 (Rel. 25, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 2(I) chain precursor.
GN COL1A2 OR COLA2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Calvaria;
RX MEDLINE=92372043; PubMed=1505972;
RA Phillips C.L., Morgan A.L., Lever L.W., Wenstrup R.J.;
RT "Sequence analysis of a full-length cDNA for the murine pro alpha
RT 2(I) collagen chain: comparison of the derived primary structure with
RT human pro alpha 2(I) collagen.";
RL Genomics 13:1345-1346 (1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Mammary gland;
RX MEDLINE=92388257; PubMed=1247932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.W., Schaefer C.F., Bhat N.K.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Hsieh F.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Sapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green B.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [3]
RP SEQUENCE OF 1-110 FROM N.A.
RC TISSUE=Calvaria;
RX MEDLINE=92084969; PubMed=1748823;
RA Phillips C.L., Lever L.W., Pinnell S.R., Quarles L.D.,
RA Wenstrup R.J.;
RT "Construction of a full-length murine pro alpha 2(I) collagen cDNA by
RT the polymerase chain reaction.";
RL J. Invest. Dermatol. 97:980-984 (1991).
RN [4]
RP SEQUENCE OF 1-23 FROM N.A.
RX MEDLINE=87289650; PubMed=3039494;
RA Rossi P., de Crombrughe B.;
RT "Identification of a cell-specific transcriptional enhancer in the
RT first intron of the mouse alpha 2 (type I) collagen gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:5590-5594 (1987).
CC -!- FUNCTION: Type I collagen is a member of group I collagen
CC (fibrillar forming collagen).
CC -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC -!- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
CC bones. In bones the fibrils are mineralized with calcium
CC hydroxyapatite.
CC -!- FTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.

```


RT 1(III)CB9B (positions 928-1028).";
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:861-868(1979).
 CC -!- FUNCTION: Collagen type III occurs in most soft connective tissues
 CC along with type I collagen.
 CC -!- SUBUNIT: Trimers of identical alpha 1(III) chains. The chains are
 CC linked to each other by interchain disulfide bonds. Trimers are
 CC also cross-linked via hydroxylysines.
 CC -!- PTM: Prolines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC PIR: A02862; CGB075.
 DR InterPro: IPR008161; Clq helix.
 DR InterPro: IPR008160; Collagen.
 DR InterPro: IPR001007; VWF_C.
 DR Pfam: PF01391; Collagen; 17.
 DR PRODOM: PD000007; Clq helix; 3.
 DR PROSITE: PS01208; VWF_C_1; PARTIAL.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen.
 FT DOMAIN 1 14 NONHELICAL REGION (N-TERMINAL).
 FT DOMAIN 15 1040 TRIPLE-HELICAL REGION.
 FT MOD RES 1041 1049 NONHELICAL REGION (C-TERMINAL).
 FT MOD RES 95 95 HYDROXYLATION.
 FT MOD RES 107 107 HYDROXYLATION.
 FT MOD RES 119 119 HYDROXYLATION.
 FT MOD RES 938 938 HYDROXYLATION.
 FT MOD RES 950 950 HYDROXYLATION.
 FT MOD RES 107 107 O-LINKED (GAL. . .).
 FT CARBOHYD 950 950 O-LINKED (GAL. . .).
 FT DISULFID 1040 1040 INTERCHAIN.
 FT DISULFID 1041 1041 INTERCHAIN.
 SQ SEQUENCE 1049 AA; 93651 MW; 8EEC33D1C66EC9A3 CRC64;
 Query Match 62.1%; Score 1905.5; DB 1; Length 1049;
 Best Local Similarity 62.7%; Pred. No. 9.1e-71;
 Matches 347; Conservative 36; Mismatches 161; Indels 9; Gaps 1;
 QY 1 GSEGEVGRGPPGPPAGAGAGADPCADGEPGAKGADGAPGAGPPGARGPSGPE 60
 DB 204 GSSGAPGQRPQGAGAGAPGPPGSDSGPGKGMGAPGAPGLIGARGPPGPP 263
 QY 61 GPGPPGPKGDSGPPGAGSGKDTGAKGPPGPPGPPGPPGPPGPPGPPGPPG 120
 DB 264 GTNGVPGQGAAGPPGKNGAKGDPGPPGPPGPPGPPGPPGPPGPPGPPG 323
 QY 121 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 180
 DB 324 GAAGGVPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 383
 QY 181 GSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 240
 DB 384 GSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 443
 QY 241 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 300
 DB 444 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 503
 QY 301 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 360
 DB 504 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 563
 QY 361 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 420
 DB 564 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 623
 QY 421 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 480
 DB 624 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 683
 QY 481 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 534
 DB 684 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 743
 QY 535 ---GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 544

DB 744 GKGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 756
 RESULT 15
 CAL3_HUMAN STANDARD; PRT; 1466 AA.
 ID F02461; Q15112;
 AC 21-JUL-1986 (Rel. 01, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Collagen alpha 1(III) chain precursor.
 GN COL3A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Skin fibroblast;
 RX MEDLINE=89350838; PubMed=2764886;
 RA Ala-Kokko L., Kontusaari S., Baldwin C.T., Kuivaniemi H.,
 RA Prockop D.J.;
 RT "Structure of cDNA clones coding for the entire prepro alpha 1 (III)
 RT chain of human type III procollagen. Differences in protein structure
 RT from type I procollagen and conservation of codon preferences.";
 RL Biochem. J. 260:509-516(1989).
 RN [2]
 RP SEQUENCE OF 149-1225 FROM N.A.
 RX MEDLINE=89386015; PubMed=2780304;
 RA Janeczko R.A., Ramirez F.;
 RT "Nucleotide and amino acid sequences of the entire human alpha 1
 RT (III) collagen.";
 RL Nucleic Acids Res. 17:6742-6742(1989).
 RN [3]
 RP SEQUENCE OF 169-398.
 RX MEDLINE=77134724; PubMed=557335;
 RA Seyer J.M., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of cyanogen
 RT bromide peptides from the amino-terminal segment of type III collagen
 RT of human liver.";
 RL Biochemistry 16:1158-1164(1977).
 RN [4]
 RP REVISIONS.
 RA Seyer J.M.;
 RL Submitted (DEC-1977) to the PIR data bank.
 RN [5]
 RP SEQUENCE OF 399-727.
 RX MEDLINE=79000343; PubMed=687591;
 RA Seyer J.M., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of five
 RT consecutive CNBr peptides from type III collagen of human liver.";
 RL Biochemistry 17:3404-3411(1978).
 RN [6]
 RP SEQUENCE OF 728-964.
 RX MEDLINE=80198282; PubMed=6246925;
 RA Seyer J.M., Mainardi C., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of alpha 1
 RT (III)-CB5 from type III collagen of human liver.";
 RL Biochemistry 19:1583-1589(1980).
 RN [7]
 RP SEQUENCE OF 950-1466 FROM N.A.
 RX MEDLINE=88189827; PubMed=3357782;
 RA Mankoo B.S., Dalgleish R.;
 RT "Human pro alpha 1(III) collagen: cDNA sequence for the 3' end.";
 RL Nucleic Acids Res. 16:2337-2337(1988).
 RN [8]
 RP REVISION TO 1184.
 RX MEDLINE=89098346; PubMed=3211760;
 RA Molyneux K., Dalgleish R.;
 RT "Human type III collagen 'variant' is a cDNA cloning artefact.";
 RL Nucleic Acids Res. 16:11833-11833(1988).
 RN [9]

VARIANT AORTIC ANEURYSM ARG-786.
MEDLINE=91056145; PubMed=2243125;
Kontusari S., Tromp G., Kuivaniemi H., Romanic A.M., Prockop D.J.;
"A mutation in the gene for type III procollagen (COL3A1) in a family
with aortic aneurysms";
J. Clin. Invest. 86:1465-1473 (1990).
[19]
VARIANT EDS-IV ARG-828.
MEDLINE=94016385; PubMed=8411057;
Richards A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.;
"The substitution of glycine 661 by arginine in type III collagen
produces mutant molecules with different thermal stabilities and
causes Ehlers-Danlos syndrome type IV";
J. Med. Genet. 30:690-693 (1993).
[20]
VARIANT EDS-IV SER-957.
MEDLINE=89109135; PubMed=2492273;
Tromp G., Kuivaniemi H., Shikata H., Prockop D.J.;
"A single base mutation that substitutes serine for glycine 790 of
the alpha 1 (III) chain of type III procollagen exposes an arginine
and causes Ehlers-Danlos syndrome IV";
J. Biol. Chem. 264:1349-1352 (1989).
[21]
VARIANT EDS-IV VAL-960.
MEDLINE=95268429; PubMed=7749417;
Tromp G., de Paeppe A., Nuytinc L., Madhathari S.L., Kuivaniemi H.;
"Substitution of valine for glycine 793 in type III procollagen in
Ehlers-Danlos syndrome type IV";
Hum. Mutat. 5:179-181 (1995).
[22]
VARIANT EDS-IV GLU-1014.
MEDLINE=92316511; PubMed=1352273;
Richards A.J., Ward P.N., Narcisi P., Nicholls A.C., Lloyd J.C.,
Pope F.M.;
"A single base mutation in the gene for type III collagen (COL3A1)
converts glycine 847 to glutamic acid in a family with Ehlers-Danlos
syndrome type IV. An unaffected family member is mosaic for the
mutation";
Hum. Genet. 89:414-418 (1992).
[23]
VARIANT EDS-IV ASP-1050.
MEDLINE=90037070; PubMed=2808425;
Tromp G., Kuivaniemi H., Stolle C.A., Pope F.M., Prockop D.J.;
"Single base mutation in the type III procollagen gene that converts
the codon for glycine 883 to aspartate in a mild variant of
Ehlers-Danlos syndrome IV";
J. Biol. Chem. 264:19313-19317 (1989).
[24]
VARIANT EDS-IV VAL-1077.
MEDLINE=91374480; PubMed=1895316;
Richards A.J., Lloyd J.C., Ward P.N., de Paeppe A., Narcisi P.,
Pope F.M.;
"Characterisation of a glycine to valine substitution at amino acid
position 910 of the triple helical region of type III collagen in a
patient with Ehlers-Danlos syndrome type IV";
J. Med. Genet. 28:458-463 (1991).
[25]
VARIANT EDS-IV GLU-1173.
MEDLINE=93022543; PubMed=1357232;
Johnson P.H., Richards A.J., Pope F.M., Hopkinson D.A.;
Query Match 61.9%; Score 1901; DB 1; Length 1466;
Best Local Similarity 63.1%; Pred. No. 1.7e-70;
Matches 345; Conservative 37; Mismatches 159; Indels. 6; Gaps 2;
1 GSEPEGVREPGPPGAGAGDPGADGCEPCAKGADGAPGTAGPFGARGPSGPE 60
360 GSNAGPQQRGEPGQGHAGACGPPPEPPGINSPPGKGMGPAGIPGAPGLMGARGPPGPA 419
61 GPGGPPGPKGDSGEPGAPGSKGDTCAKGEPPGVGEPPGPGAGEKPGARGEPGTGLP 120
420 GAGAPGLRGAGEPGKNGAKGEPGPGERGEACIPGVPAKGBDKGSGEPGANGLP 479

Search completed: September 24, 2004, 11:07:24
Job time : 14.7264 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2004, 11:06:55 ; Search time 74.4117 Seconds
(without alignments)
2065.614 Million cell updates/sec

Title: US-10-658-989A-4
Perfect score: 3070
Sequence: 1 GSECPGVRGPPGPAGCA.....PGSPGADGPPGPPGAGKEG 544

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq_29Jan04.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2966	96.6	1057	3	AAy84541 Amino aci
2	2966	96.6	1057	3	AAy84544 A human c
3	2966	96.6	1058	3	AAy84403 Amino aci
4	2966	96.6	1107	2	AAr89472 Collagen/
5	2966	96.6	1107	3	AAy84540 Amino aci
6	2966	96.6	1161	7	AAr89469 Collagen/
7	2966	96.6	1169	2	AAr89470 Collagen/
8	2966	96.6	1169	3	AAy84537 Amino aci
9	2966	96.6	1171	2	AAr89470 Collagen/
10	2966	96.6	1171	3	AAy84538 Amino aci
11	2966	96.6	1211	7	AAr89471 Collagen/
12	2966	96.6	1226	7	AAr89471 Collagen/
13	2966	96.6	1388	2	AAr89471 Collagen/
14	2966	96.6	1411	3	AAy56800 Human pre
15	2966	96.6	1461	5	ABg93947 Human pro
16	2966	96.6	1464	2	AAw68485 Human rec
17	2966	96.6	1464	4	AAw68485 Human rec
18	2966	96.6	1464	4	AAU1136 Human nov
19	2966	96.6	1464	5	ABg90764 Human tum
20	2966	96.6	1464	5	ABP68610 Human tum
21	2966	96.6	1464	6	ABU54471 Human tum
22	2966	96.6	1464	6	ABR47417 Breast ca
23	2966	96.6	1464	6	ABr92064 Human cer
24	2966	96.6	1464	6	ADD14142 Human src
25	2966	96.6	1464	7	ADD45059 Human Pro

RESULT 1
AAy84541
ID AAy84541 standard; protein; 1057 AA.
XX AC AAy84541;
XX AC
DT 25-JUL-2000 (first entry)
XX DE Amino acid sequence of a human collagen 1 (alpha) protein.
XX KW Extracellular matrix protein; self aggregation; hydroxylated proline;
trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
collagen; fibrinogen; fibronectin; post translational hydroxylation.
XX OS Homo sapiens.
XX PN EP992586-A2.
XX PD 12-APR-2000.
XX PF 07-OCT-1999; 99EP-00119184.
XX PR 09-OCT-1998; 98US-00169768.
(USSU) US SURGICAL CORP.
Gruskin EA, Buechter DD, Zhang G, Connolly K;
WPI; 2000-259138/23.
N-PSDB; AAA12502.
Production of extracellular matrix proteins containing 4-trans-
hydroxyproline results in native self aggregating proteins, useful on
medical implants.
Disclosure; Fig 27A-E; 260pp; English.
The specification describes a method for producing an extracellular
matrix protein or its fragment. The extracellular matrix protein is
capable of self aggregating in a cell which does not ordinarily
hydroxylated prolines. The method comprises optimising a nucleic acid
sequence for expression in the cell by substitution of codons preferred
by that cell for naturally occurring codons not preferred by the cell;
incorporating the nucleic acid sequence into the cell; and contacting the
cell with a hypertonic growth medium containing at least one amino acid,
selected from the group consisting of trans-4-hydroxyproline and 3-
hydroxyproline to allow at least one of the amino acids to be assimilated
into the cell and incorporated into the extracellular matrix protein. The

ALIGNMENTS

Db	561	GAPELQMPGEKGAAGLPGFKGDRGDAGPKGADGSPGKDGVRGUTGFIGPPGAGAPGDK	620
Qy	421	GESGSPGAPGPTGARGAPGDRGEPGPPGAFAGFPAGADGEPGAKGFEFGDAGAKGDAGPP	480
Db	621	GESGSPGAPGPTGARGAPGDRGEPGPPGAFAGFPAGADGEPGAKGFEFGDAGAKGDAGPP	680
Qy	481	GPAGPAGPPGPTGDTGARGAKGARGAGSAGPPGATGTFPGNAGRVGPPGSGDAGPSPGPGA	540
Db	681	GPAGPAGPPGPTGDTGARGAKGARGAGSAGPPGATGTFPGNAGRVGPPGSGDAGPSPGPGA	740
Qy	541	GKEG 544	
Db	741	GKEG 744	
RESULT 3			
AA	AA584403	ID AAY84403 standard; protein; 1058 AA.	
AC	XX	AA584403;	
XX	XX	12-JUL-2000 (first entry)	
DT	XX	Amino acid sequence of human type 1 (alpha1) collagen polypeptide.	
DE	XX	Alphal collagen; 3,4-dehydro-L-proline; epoxidation; 3,4-epoxyproline;	
KW	XX	collagen; mussel adhesive protein; bioadhesive.	
KW	XX	Homo sapiens.	
OS	XX	WO200014201-A1.	
PN	XX	16-MAR-2000.	
PD	XX	07-SEP-1999; 99WO-US020462.	
PF	XX	03-SEP-1998; 98US-0099652P.	
PR	XX	(USSU) US SURGICAL CORP.	
PA	PA	(PAOL/) PAOLELLA D N.	
PA	PA	(GRUS/) GRUSKIN E A.	
PA	PA	(BUEC/) BUECHTER D D.	
XX	XX	Paolella DN, Gruskin EA, Buechter DD;	
PI	XX	WPI; 2000-271051/23.	
DR	DR	N-PSDB; AAZ99843.	
DR	XX	Incorporating non-natural amino acid into polypeptide, useful e.g. for	
PT	PT	production of bioadhesives, by epoxidation or substitution of	
PT	PT	dehydroproline residues.	
XX	XX	Disclosure; Fig 6; 66pp; English.	
PS	XX	The present sequence represents a human type 1 (alpha1) collagen protein.	
CC	CC	Peptides derived from the protein were used to demonstrate incorporation	
CC	CC	of 3,4-dehydro-L-proline into the peptide, using the method of the	
CC	CC	invention. The specification describes a method for the incorporation of	
CC	CC	non-natural amino acid into a polypeptide. The method comprises reacting	
CC	CC	at least one 3,4-dehydroproline residue in the polypeptide with an	
CC	CC	epoxidation reagent from a polypeptide containing at least one 3,4-	
CC	CC	epoxyproline residue. The method is used for studying the effects of non-	
CC	CC	natural amino acids on structure and function of polypeptides. The method	
CC	CC	is also useful for commercial production of collagen or mussel adhesive	
CC	CC	proteins (which are useful as bioadhesives), and for incorporating a wide	
CC	CC	variety of groups, including therapeutic ligands and biological probes,	
CC	CC	into polypeptides	
XX	XX	Sequence 1058 AA;	
XX	XX	Query Match 96.6%; Score 2566; DB 3; Length 1058;	
XX	XX	Best Local Similarity 95.6%; Pred No. 2.7e-169;	

Best Local Similarity		95.6%;	Pred. No. 2.8e-159;		
Matches 520;		Conservative 21;	Mismatches 3;	Indels	0; Gaps 0;
QY	1	GSEGEVGRGEPGPPGAGAGPADGADCEPKAKGADGAPGTAGAPFPFCARGPSQPE	60		
Db	201	GSESPQGVGRGEPGPPGAGAGPADGADQCPKAKGANGAPGTAGAPFPFGARGSPSQ	260		
QY	61	GPGGPPPGKDSGEPGAPGSGKDTGAKGEPGPGVEGPPGACGREGKPGARGBPPTGLP	120		
Db	261	GPGGPPPGKNSGEPGAPGSGKDTGAKGEPGPGVQQPPGPGAGEGKRKGARGEPPTGLP	320		
QY	121	GPPEERGPGSGRFPFGADGVAGPKGAPAGERGSPGAPGKSPGAGRPGEAGLPCAKGLT	180		
Db	321	GPPEERGPGSGRFPFGADGVAGPKGAPAGERGSPGAPGKSPGAGRPGEAGLPCAKGLT	380		
QY	181	GSTGSPGPDCKTGPPGAGEDGRPGPPPCARGAEAGVMGPPGPKGAAGEPCKAGERVP	240		
Db	381	GSPGSPGPDCKTGPPGAGDGRPGPPPGARQQAQVMGPPGPKGAAGEPCKAGERVP	440		
QY	241	GPFGAVGPAKDKGEAGAEAGPPGPPGAPAGERGEEGFPAGSPGEGLPFGAGPPGCEAGKPGEE	300		
Db	441	GPFGAVGPAKDKGEAGAEAGPPGPPGAPAGERGEEGFPAGSPGEGLPFGAGPPGCEAGKPGEE	500		
QY	301	GVPGDGLCAPSPSGARGPFPFGERGVPGPCPAGPPGADGAPGDDGAKGADGAPGAPGSE	360		
Db	501	GVPGDGLCAPSPSGARGRFPFGERGVPGPCPAGFPRGANGAPGNDGAKGADGAPGAPGSG	560		
QY	361	GAPLEGMPERGAAAGLPFGKGRDAGPKGADGSPGKGVRLGTGTGIPGPAGAPGADK	420		
Db	561	GAPCLQMPERGAAAGLPFGKGRDAGPKGADGSPGKGVRLGTGTGIPGPAGAPGADK	620		
QY	421	GESGPSGAPGPTGARGAPGDRGEPGPPGPFAGFPAGDGEFGAKGFPFGDAGAKGDAGPP	480		
Db	621	GESGPSGAPGPTGARGAPGDRGEPGPPGPFAGFPAGDGPQPGAKGFPFGDAGAKGDAGPP	680		
QY	481	GPAGPAGPPGPIGDNVGAAPGAKGARGSPGATGFPGAAGRVGPPGSGDAGPPGPGPA	540		
Db	681	GPAGPAGPPGPIGNVGAAPGAKGARGSPGATGFPGAAGRVGPPGSGDAGPPGPGPA	740		
QY	541	GKEG 544			
Db	741	GKEG 744			

RESULT 6	
ADDE87050	
ID	ADDE87050 standard; protein; 1161 AA.
XX	
XX	
ADDE87050;	
XX	
AC	
AC	
XX	
DT	29-JAN-2004 (first entry)
XX	
DE	Human pancreatic cell protein sequence SegID510.
XX	
KW	neoplastic pancreatic cell; pancreatic cell; pancreatic cancer;
KW	cancer death; cytostatic; vaccine; gene therapy;
KW	non-cancerous pancreas disease; human.
XX	
OS	Homo sapiens.
XX	
PN	WO2003060145-A2.
XX	
PD	24-JUL-2003.
XX	
PF	19-DEC-2002; 2002WO-US040655.
XX	
PR	21-DEC-2001; 2001US-0342768P.
XX	
PA	(DIAD-) DIADEXUS INC.
XX	
PI	Sun Y, Liu C;
XX	
DR	WPI; 2003-587286/55.

DR N-PSDB; ADE87387.
XX
XX New pancreatic specific nucleic acid molecule or protein for diagnosing,
PT staging, imaging, monitoring, preventing or treating pancreatic cancer or
PT non-cancerous disease states of the pancreas.
XX
XX Claim 12; SEQ ID NO 510; 635pp; English.
PS
XX
XX This invention relates to novel nucleic acids and proteins present in
CC normal and neoplastic pancreatic cells. Pancreatic cancer is a common
CC cause of cancer death worldwide, therefore accurate methods of diagnosis
CC and treatment are required. Compounds which modulate the proteins of the
CC invention may have cytostatic activity and the protein and DNA sequences
CC of the invention may be useful for the development of a vaccine or in
CC gene therapy. The composition and methods are useful in diagnosing,
CC staging, imaging, monitoring, preventing or treating pancreatic cancer
CC and non-cancerous disease states of the pancreas. The present sequence is
CC that of a human pancreatic protein of the invention.
XX
XX Sequence 1161 AA;
SQ

Query Match	96.6%	Score 2966	DB 7	Length 1161
Best Local Similarity	95.6%	Pred. No. 2.9e-169		
Matches 520	Conservative 21	Mismatches 3	Indels 0	Gaps 0
Qy	1	GSEGEYGRGPPGPPGAGAAAGADPGADGCEPGAKGADGAPGIAGAPFPFGAGRGS	GP 60	
Db	59	GSEGPQVKGPPGPPGCPAGAAGPAGNPGADQCPGAKGANGAPGIAGAPFPFGAGRGS	GP 118	
Qy	61	GGGPPGPKGDSGPGAGSGKDTGAKGEPVVCVEGPPGAPAGEYKPGNAGRPCTGLP	120	
Db	119	GGGPPGPKGNSGEGAGSGKDTGAKGEPVGVQGPFGAPAGEYKRGARGEPGPTGLP	178	
Qy	121	GPPGGRGGSGRPPGADGVAGPKPAGERGSPGAPGKSGPGEAGRPGEAGLP	GA 180	
Db	179	GGPGRGGSGRPPGADGVAGPKPAGERGSGPPGAPGKSGPGEAGRPGEAGLP	GA 238	
Qy	181	GSPGSGPDGDKTGGPPGAGEDRGPPGPPGARGGAGVGMGFPGPKGAAAGEPGKAGRGVP	240	
Db	239	GSPGSGPDGDKTGGPPGAGQDGRGPPGPPGARGGAGVGMGFPGPKGAAAGEPGKAGRGVP	298	
Qy	241	GPPGAVGAGKDGAGAGBPPGPAPGAGERGEGPAGSPGFEGLPGPAGPPGAGKPGBE	300	
Db	299	GPPGAVGAGKDGAGAGBPPGPAPGAGERGEGPAGSPGFEGLPGPAGPPGAGKPGBE	358	
Qy	301	GVPGDLGAPSGARGEPFGPGRGVSGPPGPPGADGAPDDGAKGDAGAPGAPGSE	360	
Db	359	GVPGDLGAPSGARGERGFPGERGVQPPGPPGPRGANGAPCNDGAKDAGAPGSPQ	418	
Qy	361	GAPGLEHVGREGAAGLPCPKDGRDAGPKGADGSPKQCVRLGTGPIGPPGAPGAPGX	420	
Db	419	GAPGLQMPGREGAAGLPCPKDGRDAGPKGADGSPKQCVRLGTGPIGPPGAPGAPGX	478	
Qy	421	GESGSPGAGTARGAPDGRGEPFGPPGAPGADGEPGAKGEPDAGAKGDAGPP	480	
Db	479	GESGSPGAGTARGAPDGRGEPFGPPGAPGADGEPGAKGEPDAGAKGDAGPP	538	
Qy	481	FGAPGAPGPGTGDVGPAGKARGSGAPGATGPPCAGRGVCPGPGSDAGPPGPPGA	540	
Db	539	FGAPGAPGPGTIGNVAPGAKARGSGAPGATGPPCAGRGVCPGPGSDAGPPGPPGA	598	
Qy	541	GREG 544		
Db	599	GREG 602		

RESULT 7
AAR89469
ID AAR89469 standard; protein; 1169 AA.
XX
AC AAR89469;
XX
DT 01-OCT-1996 (first entry)

Db 321 GPPGEGGSGRFPFGADGVAGPKGPGAGERSGPGAGPKGSGPCEAGRPGAGLPGAKGLT 380
 Qy 181 GSPGSGPDKTGPAGDGRPGPPGPGARGEAGVGMFPKGAAGEPKKAGRGVP 240
 Db 381 GSPGSGPDKTGPAGDGRPGPPGPGARGEAGVGMFPKGAAGEPKKAGRGVP 440
 Qy 241 GPGAVGPAKDEAGAGGPPGAGPAGERGEGPAGSGPFGELPGPAGPPEAGKPGSE 300
 Db 441 GPGAVGPAKDEAGAGGPPGAGPAGERGEGPAGSGPFGELPGPAGPPEAGKPGSE 500
 Qy 301 GVPDGLGAPGSGARGEPPGPGRGVEGPPGAGPGADGAGDGAAGDAGAPGSE 360
 Db 501 GVPDGLGAPGSGARGEPPGPGRGVEGPPGAGPGADGAGDGAAGDAGAPGSE 560
 Qy 361 GAPLEGMPGEGGAAGLPGPKDGRDAGPKGADGSPKGVRLTGPFGPAGAPGDK 420
 Db 561 GAPLEGMPGEGGAAGLPGPKDGRDAGPKGADGSPKGVRLTGPFGPAGAPGDK 620
 Qy 421 GESGSPGAGTARGAGPDRGPPGPPGAGPAGPAGPAGPAGPAGPAGPAGPAGP 480
 Db 621 GESGSPGAGTARGAGPDRGPPGPPGAGPAGPAGPAGPAGPAGPAGPAGPAGP 680
 Qy 481 GPAGPAGPFPIDGVGAPGAKGARGSGPPGATGFFCAAGRVGPPGSGDAGPFP 540
 Db 681 GPAGPAGPFPIDGVGAPGAKGARGSGPPGATGFFCAAGRVGPPGSGDAGPFP 740
 Qy 541 GKEG 544
 Db 741 GKEG 744

RESULT 10
 AAY84538
 ID AAY84538 standard; protein; 1171 AA.
 AC AAY84538;
 XX 25-JUL-2000 (first entry)
 DE A chimeric collagen 1 (alpha1)/TGF-beta1 protein.
 KW Extracellular matrix protein; self aggregation; hydroxylated proline;
 KW trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
 KW collagen; fibronectin; fibronectin; post translational hydroxylation;
 KW ss. transforming growth factor-beta1; TGF-beta1; chimera.
 XX Homo sapiens.
 OS Unidentified.
 OS Chimeric.
 XX Key Location/Qualifiers
 FH Misc-difference 858
 FT /note= "Gly encoded by GCT"
 FT
 XX EF992556-A2.
 XX 12-APR-2000.
 XX 07-OCT-1999; 99EP-00119184.
 XX 09-OCT-1998; 98US-00169768.
 XX (USSU) US SURGICAL CORP.
 XX Gruskin BA, Buechter DD, Zhang G, Connolly K;
 XX WPI; 2000-259138/23.
 XX N-PSDB; AAA12498.
 XX Production of extracellular matrix proteins containing 4-trans-
 XX hydroxyproline results in native self aggregating proteins, useful on
 XX medical implants.

PS Claim 23; Fig 15; 260pp; English.
 XX The specification describes a method for producing an extracellular
 CC matrix protein or its fragment. The extracellular matrix protein is
 CC capable of self aggregating in a cell which does not ordinarily
 CC hydroxylated prolines. The method comprises optimising a nucleic acid
 CC sequence for expression in the cell by substitution of codons preferred
 CC by that cell for naturally occurring codons not preferred by the cell;
 CC incorporating the nucleic acid sequence into the cell; and contacting the
 CC cell with a hypertonic growth medium containing at least one amino acid,
 CC selected from the group consisting of trans-4-hydroxyproline and 3-
 CC hydroxyproline to allow at least one of the amino acids to be assimilated
 CC into the cell and incorporated into the extracellular matrix protein. The
 CC method may be used to make host cells assimilate and incorporate trans-4-
 CC hydroxyproline into proteins. This is especially useful in the
 CC recombinant production of proteins such as collagen, fibrinogen and
 CC fibronectin whose ability to self aggregate and produce functional
 CC proteins depends on the post translational hydroxylation of proline. The
 CC method is also useful in studying the structure and function of
 CC polypeptides which do not normally contain trans-4-hydroxyproline. The
 CC present sequence represents chimeric collagen 1 (alpha1)/transforming the
 CC growth factor-beta1 (TGF-beta1) protein, which may be produced using the
 CC method of the invention
 XX
 SQ Sequence 1171 AA;

Query Match 95.6%; Score 2966; DB 3; Length 1171;
 Best Local Similarity 95.6%; Pred. No. 2.9e-169;
 Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;
 Qy 1 GSEGPGEVGRGPPGPPGAGAGPAGDPGADGEGPAGKAGDAGPAGPFGAGSGPE 60
 Db 201 GSEGPGEVGRGPPGPPGAGAGPAGDPGADGEGPAGKAGDAGPAGPFGAGSGPE 260
 Qy 61 GPGGPPGKDSGPPGAPGSKGDTGAKGEPGVPVGEPPGPPAGGEPKAGRGPGTGLP 120
 Db 261 GPGGPPGKDSGPPGAPGSKGDTGAKGEPGVPVGEPPGPPAGGEPKAGRGPGTGLP 320
 Qy 121 GPPGERGSGSRGPPGADGVAGPKGACERGSPPGAPGKSGPCEAGRPGAGLPGAKGLT 180
 Db 321 GPPGERGSGSRGPPGADGVAGPKGACERGSPPGAPGKSGPCEAGRPGAGLPGAKGLT 380
 Qy 181 GSPGSGPDKTGPAGDGRPGPPGPGARGEAGVGMFPKGAAGEPKKAGRGVP 240
 Db 381 GSPGSGPDKTGPAGDGRPGPPGPGARGEAGVGMFPKGAAGEPKKAGRGVP 440
 Qy 241 GPPGAVGPAKDEAGAGGPPGAGPAGERGEGPAGSGPFGELPGPAGPPEAGKPGSE 300
 Db 441 GPPGAVGPAKDEAGAGGPPGAGPAGERGEGPAGSGPFGELPGPAGPPEAGKPGSE 500
 Qy 301 GVPDGLGAPGSGARGEPPGPGRGVEGPPGAGPGADGAGDGAAGDAGAPGSE 360
 Db 501 GVPDGLGAPGSGARGEPPGPGRGVEGPPGAGPGADGAGDGAAGDAGAPGSE 560
 Qy 361 GAPLEGMPGEGGAAGLPGPKDGRDAGPKGADGSPKGVRLTGPFGPAGAPGDK 420
 Db 561 GAPLEGMPGEGGAAGLPGPKDGRDAGPKGADGSPKGVRLTGPFGPAGAPGDK 620
 Qy 421 GESGSPGAGTARGAGPDRGPPGPPGAGPAGPAGPAGPAGPAGPAGPAGPAGP 480
 Db 621 GESGSPGAGTARGAGPDRGPPGPPGAGPAGPAGPAGPAGPAGPAGPAGPAGP 680
 Qy 481 GPAGPAGPFPIDGVGAPGAKGARGSGPPGATGFFCAAGRVGPPGSGDAGPFP 540
 Db 681 GPAGPAGPFPIDGVGAPGAKGARGSGPPGATGFFCAAGRVGPPGSGDAGPFP 740
 Qy 541 GKEG 544
 Db 741 GKEG 744

AD87057	standard; protein; 1211 AA.
AD87057	
29-JAN-2004	(first entry)
Human	pancreatic cell protein sequence SeqID517.
neoplastic	pancreatic cell; pancreatic cell; pancreatic cancer;
cancer death;	cytostatic; vaccine; gene therapy;
non-cancerous	pancreas disease; human.
Homo sapiens.	
WC2003060145-A2.	
24-JUL-2003.	
19-DEC-2002;	2002WO-US040655.
21-DEC-2001;	2001US-0342768P.
(DIAD-) DIADEXUS INC.	
Sun Y, Liu C;	
WPI; 2003-587286/55.	
N-PSDB; AD887397.	
New pancreatic	specific nucleic acid molecule or protein for diagnosing,
staging, imaging,	monitoring, preventing or treating pancreatic cancer or
non-cancerous	disease states of the pancreas.
Claim 12;	SEQ ID NO 517; 635pp; English.
This invention	relates to novel nucleic acids and proteins present in
normal and	neoplastic pancreatic cells. Pancreatic cancer is a common
cause of cancer	death worldwide, therefore accurate methods of diagnosis
and treatment	are required. Compounds which modulate the proteins of the
invention may	have cytostatic activity and the protein and DNA sequences
of the invention	may be useful for the development of a vaccine or in
gene therapy.	The composition and methods are useful in diagnosing,
staging, imaging,	monitoring, preventing or treating pancreatic cancer is
and non-cancerous	disease states of the pancreas. The present sequence is
that of a human	pancreatic protein of the invention.
Sequence 1211	AA;
Query Match	96.6%; Score 2966; DB 7; Length 1211;
Best Local Similarity	95.6%; Pred. No. 3e-169;
Matches 520;	Conservative 21; Mismatches 3; Indels 0; Gaps 0;
1	GSEGEVGRGEPGPPGAGAGPAGDPCADGEPGAKGADGAPGIAGAPGPPGARGPSGPE 60
362	GSEGPQGVGRGEPGPPGAGAGPAGNPGADQPGKANGAPGIAGAPGPPGARGPSGPQ 421
61	GGCGPPGKDGSEFGARGSGKDTGAKGEPGVGEPGPPGAGEGKPGARGEPGTGLP 120
422	GGPGPPGPKNGSEGFAPGSGKDTGAKGEPGVGQPGPPGAGEGKRGARGEPGTGLP 481
121	GPPEGRGGSGSFFGADGVAGPKPAGERSGPPAGPKGSPGAEGRGEPGAGLPKAGL 180
482	GPPEGRGGSGSFFGADGVAGPKPAGERSGPPAGPKGSPGAEGRGEPGAGLPKAGL 541
181	GGPSGPPGKGTGPPGAGEGRGPPGPPGARGGAGVMGFPQPKGAAEGEPGKAGRGVP 240
542	GGPSGPPGKGTGPPGAGQDGRGPPGPPGARGQAGVMGFPQKGAEGEPGKAGRGVP 601
241	GGPGAVGPKDGEAGAGSPGPPGAGEGEGPAGSGPFGELPGPAGPPGEGAGKPGEE 300
602	GGPGAVGPKDGEAGAGSPGPPGAGEGEGPAGSGPFGELPGPAGPPGEGAGKPGEE 360
301	GVFDGLGAPGSPGARGEGEFGFGEGRGVEGPPGAGPPGADGAGDGDGKADGAPGAGE 360

Db	662	GVFGDLGAPGSGGARGGFFGGKGVGGPPGAPGPRGANGAPGNGCAKGDGAGAPGPGSQ	721
Qy	361	GAPGLEGMFGBERGAAGLFGPKGDRGDAGPKGADGSPCKDGVRLGTGPIGPPGAPGAPGDK	420
Db	722	GAPGLQGMFGBERGAAGLFGPKGDRGDAGPKGADGSPCKDGVRLGTGPIGPPGAPGAPGDK	781
Qy	421	GESGSGPAGTGAAGAGGDEGEGPPGAPGAPGAGPCADGEPCKAGCEPCAGKGDAGPP	480
Db	782	GESGSGPAGTGAAGAGGDEGEGPPGAPGAPGAGPCADGEPCKAGCEPCAGKGDAGPP	841
Qy	481	GPAGPAGPPGPIGVGAPGAKGARGSGAPPGATGFPGAAGRVGPPGSGDAGPPGPPGPA	540
Db	842	GPAGPAGPPGPIGVGAPGAKGARGSGAPPGATGFPGAAGRVGPPGSGDAGPPGPPGPA	901
Qy	541	GKEG 544	
Db	902	GKEG 905	
RESULT 12			
ADE87062			
XX	XX	ADAE87062 standard; protein; 1226 AA.	
XX	XX	ADAE87062;	
XX	XX	29-JAN-2004 (first entry)	
XX	XX	Human pancreatic cell protein sequence SeqID522.	
XX	XX	neoplastic pancreatic cell; pancreatic cancer;	
XX	XX	cancer death; cytostatic; vaccine; gene therapy;	
XX	XX	non-cancerous pancreas disease; human.	
XX	XX	Homo sapiens.	
XX	XX	WO2003060145-A2.	
XX	XX	24-JUL-2003.	
XX	XX	19-DEC-2002; 2002WO-US040655.	
XX	XX	21-DEC-2001; 2001US-0342768P.	
XX	XX	(DIAD-) DIADEXUS INC.	
XX	XX	Sun Y, Liu C;	
XX	XX	WPI; 2003-587286/55.	
XX	XX	N-PSDB; ADE87403.	
XX	XX	New pancreatic specific nucleic acid molecule or protein for diagnosing,	
XX	XX	staging, imaging, monitoring, preventing or treating pancreatic cancer or	
XX	XX	non-cancerous disease states of the pancreas.	
XX	XX	Claim 12; SEQ ID NO 522; 635pp; English.	
XX	XX	This invention relates to novel nucleic acids and proteins present in	
XX	XX	normal and neoplastic pancreatic cells. Pancreatic cancer is a common	
XX	XX	cause of cancer death worldwide, therefore accurate methods of diagnosis	
XX	XX	and treatment are required. Compounds which modulate the proteins of the	
XX	XX	invention may have cytostatic activity and the protein and DNA sequences	
XX	XX	of the invention may be useful for the development of a vaccine or in	
XX	XX	gene therapy. The composition and methods are useful in diagnosing,	
XX	XX	staging, imaging, monitoring, preventing or treating pancreatic cancer	
XX	XX	and non-cancerous disease states of the pancreas. The present sequence is	
XX	XX	that of a human pancreatic protein of the invention.	
XX	XX	Sequence 1226 AA;	
XX	XX	SQ	

Query Match	96.6%;	Score 2966;	DB 7;	Length 1226;
Best Local Similarity	95.6%;	Pred. No. 3e-169;		

XX	12-JUN-1995;	95CA-02151547.
XX	PF	
XX	XX	
PR	10-JUN-1994;	94US-00259263.
XX	XX	
XX	XX	
PA	(USSU) US SURGICAL CORP.	
XX	XX	
Pi	Gruskin EA, Espino P;	
XX	XX	
DR	WPI; 1996-140144/15.	
DR	N-PSDB; AAT16517.	
XX	XX	
XX	Chimeric DNA encoding protein contg. extracellular matrix protein domain	
PT	- and cellular regulatory factor domain, partic. useful as osteogenic	
PT	agents, also related vectors, transformed cells and chimaeric proteins.	
XX	XX	
PS	Disclosure; Fig 7; 59pp; English.	
XX	XX	
CC	A fusion protein (AA889471) comprises the alpha-helical region of human	
CC	collagen I(a) linked to human dermatan sulphate proteoglycan (decorin).	
CC	It can be expressed in Escherichia coli transformants carrying a vector	
CC	incorporating a chimeric gene (AA816517) coding for the fusion. The	
CC	decorin binds to type I collagen and thus affects Elbri1 formation. It	
CC	inhibits the cell attachment-promoting activity of collagen and	
CC	fibrinogen by binding to such molecules near their cell binding sites.	
CC	The collagen moiety provides an integral substratum or scaffolding for	
CC	the decorin. The fusion protein acts to reduce scarring of healing tissue	
XX	XX	
SQ	Sequence 1388 AA;	

Query Match	96.6%	Score 2966;	DB 2;	Length 1388;
Best Local Similarity	95.6%	Pred. No. 3.4e-159;		
Matches 550;	Conservative 21;	Mismatches 3;	Indels 0;	Gaps 320
Qy	1	GSEGPQVGRGEPGPPGAGAGADPGADGEPGAKGADGAPGIAGAPGPPGARGSPGE	60	
Db	201	GSEGPQVGRGEPGPPGAGAGAGADPGADGEPGAKGADGAPGIAGAPGPPGARGSPGE	260	
Qy	61	GGGGPPGKGGSGEPGAPGSKDGTAKGEGPQVGVGEGPGPAGGEEGKAGARGEPGTGLP	120	
Db	261	GGGGPPGKGGSGEPGAPGSKDGTAKGEGPQVGVGEGPGPAGGEEGKAGARGEPGTGLP	320	
Qy	121	GGPGERGGPGRGPPGADGVAGPKPAGERGSGPGAPGKSGPGEAGRPGEAGLPKAGKLT	180	
Db	321	GGPGERGGPGRGPPGADGVAGPKPAGERGSGPGAPGKSGPGEAGRPGEAGLPKAGKLT	380	
Qy	181	GGSPGSGPDGKTGPPGPAGEGDRGPPGPPGARGGAGVGMGFPKPKAAAGEPGKAGARGVP	240	
Db	381	GGSPGSGPDGKTGPPGPAGEGDRGPPGPPGARGGAGVGMGFPKPKAAAGEPGKAGARGVP	440	
Qy	241	GGPVAGVAGKDXGAGAGGPPGAPGAPAGERGEBGPAGSPGFEGLPGAPGPGGAGKPGSE	300	
Db	441	GGPVAGVAGKDXGAGAGGPPGAPGAPAGERGEBGPAGSPGFEGLPGAPGPGGAGKPGSE	500	
Qy	301	GVPDGLGAPSGGARGGEPFPGERGVBGPPGAPGPGADGAPDGDGAKGDAGAPAGPGE	360	
Db	501	GVPDGLGAPSGGARGGEPFPGERGVBGPPGAPGPGADGAPDGDGAKGDAGAPAGPGE	560	
Qy	361	GAPGLEHWPGBERGAAAGLPGPKDGRDAGPKAGDGS PGKQVRLGTGPIGPBPAGAPGDK	420	
Db	561	GAPGLEHWPGBERGAAAGLPGPKDGRDAGPKAGDGS PGKQVRLGTGPIGPBPAGAPGDK	620	
Qy	421	GSGSPSGAGPTGARGAPDRGEPGPPGAPGAGPAGPGADGEPGAKGEPDAGAKGDAGPP	480	
Db	621	GSGSPSGAGPTGARGAPDRGEPGPPGAPGAGPAGPGADGEPGAKGEPDAGAKGDAGPP	680	
Qy	481	GAGAPAGPBGPTGDVAPGAKARGSGAGPGATGFGAGARGVCPGPGSGDAGPGPPGPA	540	
Db	681	GAGAPAGPBGPTGDVAPGAKARGSGAGPGATGFGAGARGVCPGPGSGDAGPGPPGPA	740	
Qy	541	GKEG 544		
Db	741	GKEG 744		

AX
CC The invention discloses a method for stimulating or inhibiting cell
CC growth and/or division which comprises contacting or inserting into an
CC animal cell a polypeptide comprising one of the deer antler cartilage

Db

479 CBBCEBGGBCSPCEBGCAGVAGCPKCPACGEPGSPGCPACGPKYCSPCFACGPPCFACI.DCAKGI.T 538

479 CBBCEBGGBCSPCEBGCAGVAGCPKCPACGEPGSPGCPACGPKYCSPCFACGPPCFACI.DCAKGI.T 538

CC cell (DACC) clones disclosed. More particularly, the method relates to
 CC these polypeptides stimulating mesenchymal cell growth and/or division
 CC and to transfecting these cells and chondrocytes with vectors carrying
 CC the genes of these polypeptides capable of stimulating chondrogenesis,
 CC osteogenesis, growth, repair, regeneration and/or restoration of the
 CC extracellular matrix. The chondrocytes selectively express genes required
 CC to form a cartilaginous matrix. The DACC polypeptides and polynucleotides
 CC are useful for identifying an agent that modulates the activity of the
 CC polypeptide, for stimulating mesenchymal cell growth and/or division by
 CC exposing animal mesenchymal cells to conditioned media or its active
 CC fraction, obtained from deer antler cartilage cells, for inhibiting cell
 CC growth and/or division by inserting into an animal cell, a compound which
 CC inhibits the translation of the polynucleotide encoding the DACC. The
 CC method and the polypeptides are useful for stimulating mesenchymal cell
 CC growth and/or division or for stimulating chondrogenesis, cartilage, disc
 CC or connective tissue growth, repair, regeneration and/or restoration in
 CC an animal. The polynucleotides, polypeptides, agonists and antagonists
 CC may be used in treatment modalities, specifically in gene therapy. The
 CC polypeptides can be used as bait proteins in a two- or three-hybrid assay
 CC to identify other proteins, which bind to or interact with the
 CC polypeptide and are involved in modulating cell growth and/or division.
 CC The sequences presented in ABG93923-ABG93948 are the proteins encoded by
 CC the DACC cDNA clones

xx
 SQ Sequence 1461 AA;

Query Match	96.6%;	Score 2966;	DB 5;	Length 1461;
Best Local Similarity	95.6%;	Pred. No. 3.5e-169;		
Matches 520;	Conservative 21;	Mismatches 3;	Indels 0;	Gaps 0;

Qy	1	GSEGEVGRGEPGPPGAGAGDGDGDEGAKGADGAGIAGAPGPPGARGSGPE	60
Db	359	GSEGPQGVGRGEPGPPGAGAGDGDGDEGAKGADGAGIAGAPGPPGARGSGPPQ	418
Qy	61	GPGGPPGPKGDSGPPGAPGKGDGTGAKGEPGVGVEGPPGAGEGKPKGARGSPGPTGLP	120
Db	419	GPGGPPGPKGNSGPPGAPGKGDGTGAKGEPGVGVQGPAGEGKRGARGSPGPTGLP	478
Qy	121	GPPGERGGPSRGPFGADGVAGPKGAGERGSPGAPGKSGPGEAGRPGEAGLPGAKGLT	180
Db	479	GPPGERGGPSRGPFGADGVAGPKGAGERGSPGAPGKSGPGEAGRPGEAGLPGAKGLT	538
Qy	181	GSPGSPGDKGTGPPGAGEGDPGPPGAGEGDPGPPGAGEGDPGPPGAGEGDPGPPG	240
Db	539	GSPGSPGDKGTGPPGAGEGDPGPPGAGEGDPGPPGAGEGDPGPPGAGEGDPGPPG	598
Qy	241	GPPGAVGPAKDGAGAGGPPGPPGAGPAGERGEGEPGAGSPGEGLPAGPPGAGKPGEE	300
Db	599	GPPGAVGPAKDGAGAGGPPGPPGAGPAGERGEGEPGAGSPGEGLPAGPPGAGKPGEE	658
Qy	301	GVPCDLGAPGSGARGPFPFGERGVEGPPGPPGAGPPGADGADGADGADGADGADG	360
Db	659	GVPCDLGAPGSGARGPFPFGERGVEGPPGPPGAGPPGADGADGADGADGADGADG	718
Qy	361	GAPGLEQMPGERGAAGLPKPKGRDAGPKGADGSPGKGVRLTGP1GPPGPAAGPDK	420
Db	719	GAPGLEQMPGERGAAGLPKPKGRDAGPKGADGSPGKGVRLTGP1GPPGPAAGPDK	778
Qy	421	GEGSPGPAFTGARGAPGRGPPGPPGAGTGPAGTGPAGTGPAGTGPAGTGPAGTGP	480
Db	779	GEGSPGPAFTGARGAPGRGPPGPPGAGTGPAGTGPAGTGPAGTGPAGTGPAGTGP	838
Qy	481	GPAGPAGPPIGIVGAFGAKGARGAGPPGATGFPAGAGRVGPPGSPGADGPPGPPGA	540
Db	839	GPAGPAGPPIGIVGAFGAKGARGAGPPGATGFPAGAGRVGPPGSPGADGPPGPPGA	898
Qy	541	GKEG 544	
Db	899	GKEG 902	

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2004, 11:06:56 ; Search time 22.3958 Seconds
(without alignments)
1254.011 Million cell updates/sec

Title: US-10-658-989A-4

Perfect score: 3070

Sequence: 1 GSEGGVGRGPPGPPGAGA.....PGPSDAGPPGPPGAGKEG 544

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents_AA.*

- 1: /cgn2_6/ptodata/2/iaa/5A COMB.pap.*
- 2: /cgn2_6/ptodata/2/iaa/5B COMB.pap.*
- 3: /cgn2_6/ptodata/2/iaa/6A COMB.pap.*
- 4: /cgn2_6/ptodata/2/iaa/6B COMB.pap.*
- 5: /cgn2_6/ptodata/2/iaa/PTUS COMB.pap.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2966	96.6	1461	4	US-09-585-887-9
2	2966	96.6	1461	4	US-09-289-578-9
3	2966	96.6	1464	4	US-09-331-347C-21
4	2857	93.1	822	3	US-09-219-849-49
5	2842	92.6	1057	3	US-08-931-820-1
6	2791.5	90.9	1341	3	US-08-963-825-18
7	2791.5	90.9	1341	4	US-09-500-811-18
8	2791.5	90.9	1341	4	US-09-570-573-18
9	2791.5	90.9	1341	4	US-09-548-608-18
10	2226	72.5	1017	4	US-08-468-986-10
11	2202	71.7	1060	3	US-08-931-820-3
12	2202	71.7	1418	3	US-08-963-825-20
13	2202	71.7	1418	3	US-09-010-999-1
14	2202	71.7	1418	4	US-09-500-811-20
15	2202	71.7	1418	4	US-09-570-573-20
16	2202	71.7	1418	4	US-09-548-608-20
17	2192	71.4	1442	2	US-08-316-650-12
18	2192	71.4	1442	5	PCT-US95-02251-12
19	2156	70.2	595	3	US-09-219-849-48
20	2156	70.2	595	3	US-09-219-849-50
21	1901	61.9	1057	3	US-08-931-820-4
22	1886	61.4	1024	3	US-08-931-820-2
23	1886	61.4	1365	3	US-08-963-825-19
24	1886	61.4	1366	4	US-09-500-811-19
25	1886	61.4	1366	4	US-09-570-573-19
26	1886	61.4	1366	4	US-09-548-608-19
27	1885	61.4	1366	4	US-09-585-887-10

ALIGNMENTS

RESULT 1

US-09-585-887-9

; Sequence 9, Application US/09585887

; Patent No. 6413742

; GENERAL INFORMATION:

; APPLICANT: Olsen, David R

; APPLICANT: Chang, Robert

; APPLICANT: McMullin, Hugh

; APPLICANT: Hitzeman, Ronald A.

; APPLICANT: Chisholm, George

; TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND

; TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT

; TITLE OF INVENTION: CELLS

; FILE REFERENCE: 225002030400

; CURRENT APPLICATION NUMBER: US/09/585,887

; CURRENT FILING DATE: 2000-05-31

; PRIOR APPLICATION NUMBER: 09/289,578

; PRIOR FILING DATE: 1999-04-09

; PRIOR APPLICATION NUMBER: 60/084,828

; PRIOR FILING DATE: 1998-05-08

; NUMBER OF SEQ ID NOS: 11

; SOFTWARE: Patent in Ver. 2.0

; SEQ ID NO 9

; LENGTH: 1461

; TYPE: PRT

; ORGANISM: Homo sapiens

; US-09-585-887-9

Query Match 96.6%; Score 2966; DB 4; Length 1461;

Best Local Similarity 95.6%; Pred. No. 8.9e-177;

Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;

QY	1	GSEGGVGRGPPGPPGAGA	PGADGPGAGKADGAPGIAGAPGPGAGPSGPE	60
DB	359	GSEGGVGRGPPGPPGAGA	PGADGPGAGKADGAPGIAGAPGPGAGPSGPE	418
QY	61	GPGGPGPKGDSGEPGAPG	SGKDTGAKGEPGVGVGPPGPGAGGKPGGARGEPGPTGLP	120
DB	419	GPGGPGPKGDSGEPGAPG	SGKDTGAKGEPGVGVGPPGPGAGGKPGGARGEPGPTGLP	478
QY	121	GPGRGGPGSGRGFGADG	VAGPKPAGRSFGPAGPKSGPGAGRGGEAGLPGAKGLT	180
DB	479	GPGRGGPGSGRGFGADG	VAGPKPAGRSFGPAGPKSGPGAGRGGEAGLPGAKGLT	538
QY	181	GSFGSGPPGDKTGPPGP	AGDGRGPPGARGAGVWGFPGPKGAAGEPGKAGRGVP	240
DB	539	GSFGSGPPGDKTGPPGP	AGDGRGPPGARGAGVWGFPGPKGAAGEPGKAGRGVP	598
QY	241	GPFGVAGPAGKDGAGAG	EPGPPGAGRGEGPAGSGPFGGLPGPAGPPGGAOKPGBE	300

Query Match	92.6%	Score	2842	DB 3	Length	1057			
Best Local Similarity	91.9%	Pred.	No. 3.3e-169						
Matches	500	Conservative	24	Mismatches	20	Indels	0	Gaps	0
QY	1	GSSEPGVGVGEP	PPGAGAAAGPAGDPGADGEPGAKGADGAPGACAGAPGPPGARGSGSGPE	60					
Db	201	GSSEPGVGVGEP	PPGAGAGNPGADGQPSAKGANGAPGACAGAPGPPGARGSGSGPQ	260					
QY	61	GGGPPGPKGDS	GEPAQSGKBDTGAKGPPVGVGPPGPPAGEEOKPGARGGPPGPTGLP	120					
Db	261	GGPGPTCARGL	VGPPAGSKESGNKGFPSAGPQPPGPGSGEGRGNGEAGSAGPP	320					
QY	121	GPGEGRGGSGR	FFCADGVAGPKGPAGERSGPGAPKGSPCGEACRPGSAGLPFGAKGLT	180					
Db	321	GPFDLRGGSGR	FFCADGVAGPKGPAGERSGPGAPKGSPCGEACRPGSAGLPFGAKGLT	380					
QY	181	GSFGSGPPDKT	GPFGPAGEDBPPGPPGARGAGVNGFPCKAAGCPGKAGRGVP	240					
Db	381	GSFGSGPPDKT	GPFGPAGQDGRFPFPFGPKAQGVNGFPCKAAGCPGKAGRGVP	440					
QY	241	GPPGAVGPAGK	GEAGAPPPGAPGAGERGEGPAGSPGFEGLPGPAGPPGAGKPGEE	300					
Db	441	GPPGAVGPAGK	GEAGAPPPGAPGAGERGEGPAGSPGFEGLPGPAGPPGAGKPGEQ	500					

QY 301 GVPDGLGAPGSGARGPFGPFGVGGPPGADGAPGDDGAKGADGAPGSGE 360
DB 501 GVPDGLGAPGSGARGPFGPFGVGGPPGADGAPGDDGAKGADGAPGSGE 560
QY 361 GAPLGEMPGERGAAGLPGPKGDRGDAGPKGADGSPKGVRLTGPFGPPGAPGDK 420
DB 561 GAPLQMPGERGAAGLPGPKGDRGDAGPKGADGSPKGVRLTGPFGPPGAPGDK 620
QY 421 GESGSPGAPGTGARGAPGDRGPPGPPGAGFAGPPGADGEPGAKGEPDAGAKGAGPP 480
DB 621 GESGSPGAPGTGARGAPGDRGPPGPPGAGFAGPPGADGEPGAKGEPDAGAKGAGPP 680
QY 481 GPAGPAGPPGPIGNVGAAPGAKGAGSAGPPGATGFPGAAGRVGPPGSGNAGPPGPPG 540
DB 681 GPAGPAGPPGPIGNVGAAPGAKGAGSAGPPGATGFPGAAGRVGPPGSGNAGPPGPPG 740
QY 541 GKEG 544
DB 741 GKEG 744

RESULT 6

US-08-963-825-18
; Sequence 18, Application US/08963825
; Patent No. 6110689
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/963,825
; FILING DATE:
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/187,319
; FILING DATE: 21-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1341 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (1)

Query Match

90.9%; Score 2791.5; DB 3; Length 1341;

Best Local Similarity 87.6%; Pred. No. 5.5e-166;
Matches 500; Conservative 21; Mismatches 23; Indels 27; Gaps 4;
QY 1 GSEGPGEVGRGPPGPPGAGAPGADGADGE---PGAKGADGAPGAGPAGPFGAR-- 54
DB 211 GASGPMGRGPPGPPGKGBBZAGKFCRCZRGCPGPPZGARGLPGTAGLPGMKHGRGFS 270
QY 55 -----GSGSEBEG-----PG--GPPPGKDSGEPGAPGSKGDTGAKGEGPV 93
DB 271 GLBGAKGAGAPGPKGZFGSPGZBCAPGZMGPPPGKNSGEPGAPGSKGDTGAKGEGPV 330
QY 94 GVEGPPGAPGEGKPGARGBPCTGLPCGRCRGSGRPPGADGVAGPKGPAGERGSP 153
DB 331 GVQGPAGGEGKRGARGEPCTGLPCGRCRGSGRPPGADGVAGPKGPAGERGSP 390
QY 154 GPAGKSGPAGRPGEAGLPGAKGLTGSPPSPDPDKTGPAGBDRGPPGPPGAR 213
DB 391 GPAGKSGPAGRPGEAGLPGAKGLTGSPPSPDPDKTGPAGBDRGPPGPPGAR 450
QY 214 GEAGVMGPPGPKGAGEPKAGRGVPCGAVGPKGDEGAGGPPGAPGAGERGEE 273
DB 451 QGAVMGPPGPKGAGEPKAGRGVPCGAVGPKGDEGAGGPPGAPGAGERGEE 510
QY 274 GPAGSPGEGLPAGPAGPPGEAGKPEEGVPGDLGAPGFSGARGPFPFGVGGPPGPA 333
DB 511 GPAGSPGEGLPAGPAGPPGEAGKPEEGVPGDLGAPGFSGARGPFPFGVGGPPGPA 570
QY 334 GPPGADGAPDDGAKGADGAPGAPGSEGAPGEGMPCGERGAGLPGPKGDRGDAGPKGAD 393
DB 571 GPRGANGAPNDGAKGADGAPGAPGSCAPGLOQMPGERGAGLPGPKGDRGDAGPKGAD 630
QY 394 GSPGKDGVRGLTGPICPPGAPGAPGDKGSPGAPGTGARGAPGDRGPPGPPGAPGA 453
DB 631 GSPGKDGVRGLTGPICPPGAPGAPGDKGSPGAPGTGARGAPGDRGPPGPPGAPGA 690
QY 454 GPPGADGEPGAKGEPGADGAKGADGAPGPPGAPGAPGPPGICGVGAPGAKGARGAGPPGAT 513
DB 691 GPPGADGEPGAKGEPGADGAKGADGAPGPPGAPGAPGPPGICGVGAPGAKGARGAGPPGAT 750
QY 514 GPPGAGSVGPPGSGDAGPPGPPGAPGK 544
DB 751 GPPGAGSVGPPGSGNAGPPGPPGAPGK 781
RESULT 7
US-09-500-811-18
; Sequence 18, Application US/09500811
; Patent No. 632314
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/500,811
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

QY 121 GPFGERGGSGRFPAGDGVAGPKGAGRGSGFGPAGPKGSGPGEAGLPGAKGLT 180
Db 435 GPFGERGAPGNRFGPGQDGLAGPKGAPGERPSGLAGPKGANGDPGRPGPGLPGARGLT 494
QY 181 GSPGSGPDGKTPPGPAGDGRPGPPGPPGARGAGVGMFGPKKGAAGPCKAGRGVY 240
Db 495 GRPDGAGPQGVGSGAPGDPGRPPGPPGARGAGVGMFGPKKGAAGPCKAGRGVY 554
QY 241 GPPGAVGPPAGKGEAGAGPFPAGPAGRGSGPAGSGPAGSGPAGSGPAGSGPAGSGP 300
Db 555 GAFGLRLGPKDGETGAEFGPPGAGRGSGPAGSGPAGSGPAGSGPAGSGPAGSGPAGSGP 614
QY 301 GVPDGLGAPGSGARGPFGPGRGVEGPPGAPGADGADGADGADGADGADGADGADGADG 360
Db 615 GVPGEAGAPGVLGPRGSGPGRGSGPAGSGPAGSGPAGSGPAGSGPAGSGPAGSGPAGSGP 674
QY 361 GAPLEGMPGEGAGAGLPGPKGDRGADGPKGADGSGPKGDRGVLGTPGPPGAGAGGDK 420
Db 675 GPPGLQGMPPGEGAGAGLPGPKGDRGADGPKGADGSGPKGDRGVLGTPGPPGAGAGGDK 734
QY 421 GESGSPGAPGTGARGAPDRGPPGPPGAPGAPGADGEPGAKGEPGADGAGDAGPP 480
Db 735 GEVGPFPAGSAGARGAPRGERTGPPGTSIAGPPGADGQPGAKGQSGAGQKGDAGAP 794
QY 481 GPAGPAGPPGPIGVGAPGAKGARGSPGATGFPGAAGRVGPPGPPGSDAGPFPGP 540
Db 795 GPQGSAPGPPGQGTGVTGPKARGAQCPGATGFPGAAGRVGPPGPPGSDAGPFPGP 854
QY 541 GKEG 544
Db 855 GKDG 858

RESULT 14
US-09-500-811-20
; Sequence 20, Application US/09500811
; Patent No. 6323314
; GENERAL INFORMATION:
; APPLICANT: Oviast, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/500,811
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 20:

SEQUENCE CHARACTERISTICS:
; LENGTH: 1418 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN -ALPHA 1 (II)
US-09-500-811-20
Query Match 71.7%; Score 2202; DB 4; Length 1418;
Best Local Similarity 71.9%; Pred. No. 2.1e-129;
Matches 391; Conservative 40; Mismatches 113; Indels 0; Gaps 0;
QY 1 GSEGEVGRGPPGPPGAPAGADGADGADGADGADGADGADGADGADGADGADGADGADG 60
Db 315 GPEGAQGRGPPGPPGAPAGADGADGADGADGADGADGADGADGADGADGADGADGADG 374
QY 61 GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGVGVGVEGPPGAGGEGKFGARGEGPFTGLP 120
Db 375 GATGLPLGPKGTGKFIAGFKGEGQPKGEPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 434
QY 121 GPPGRRGPPGRRGPPGADGAGVAGPKGAPGAGRGSPGAPGAPGAPGAPGAPGAPGAPGAP 180
Db 435 GPPGRRGAPGNRFGPPGQDGLAGPKGAPGERGPPGSLAGPKGANGDPGRPGPGLPGARGLT 494
QY 181 GSPGSPGPDGKTGPPGPPGAGEDGGRGPPGPPGARGGAGVGMFGPPGKAAAGEPKAGRGVY 240
Db 495 GRPDGAGPQGVGSGAPGDPGRPPGPPGAPGADGADGADGADGADGADGADGADGADGADG 554
QY 241 GPPGAVGPPAGKGEAGAGPFPAGPAGRGSGPAGSGPAGSGPAGSGPAGSGPAGSGPAGSGP 300
Db 555 GAFGLRLGPKDGETGAEFGPPGAGRGSGPAGSGPAGSGPAGSGPAGSGPAGSGPAGSGPAGSGP 614
QY 301 GVPDGLGAPGSGARGPFGPGRGVEGPPGAPGADGADGADGADGADGADGADGADGADGADG 360
Db 615 GVPGEAGAPGVLGPRGSGPGRGSGPAGSGPAGSGPAGSGPAGSGPAGSGPAGSGPAGSGPAGSGP 674
QY 361 GAPLEGMPGEGAGAGLPGPKGDRGADGPKGADGSGPKGDRGVLGTPGPPGAGAGGDK 420
Db 675 GPPGLQGMPPGEGAGAGLPGPKGDRGADGPKGADGSGPKGDRGVLGTPGPPGAGAGGDK 734
QY 421 GESGSPGAPGTGARGAPDRGPPGPPGAPGAPGADGEPGAKGEPGADGAGDAGPP 480
Db 735 GEVGPFPAGSAGARGAPRGERTGPPGTSIAGPPGADGQPGAKGQSGAGQKGDAGAP 794
QY 481 GPAGPAGPPGPIGVGAPGAKGARGSPGATGFPGAAGRVGPPGPPGSDAGPFPGP 540
Db 795 GPQGSAPGPPGQGTGVTGPKARGAQCPGATGFPGAAGRVGPPGPPGSDAGPFPGP 854
QY 541 GKEG 544
Db 855 GKDG 858

RESULT 15
US-09-570-573-20
; Sequence 20, Application US/09570573
; Patent No. 6342361
; GENERAL INFORMATION:
; APPLICANT: Oviast, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York

